(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number
WO 01/21644 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/EP00/09325

(22) International Filing Date:
25 September 2000 (25.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
P 9902127 24 September 1999 (24.09.1999) ES
P 9902474 11 November 1999 (11.11.1999) ES

(71) Applicant *(for all designated States except US)*: **CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS** [ES/ES]; Calle Serrano, 117, E-28006 Madrid (ES).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GUTIERREZ-ARMENTA, Crisanto** [ES/ES]; Centro de Biología Molecular "Severo Ochoa", Universidad Autónoma, Cantoblanco, E-28049 Madrid (ES). **RAMIREZ-PARRA, Elena** [ES/ES]; Centro de Biología Molecular "Severo Ochoa", Universidad Autónoma, Cantoblanco, E-28049 Madrid (ES).

(74) Agent: **UNGRIA, Javier**; Ungría Patentes y Marcas, S.A., Avda. Ramón y Cajal, 78, E-28043 Madrid (ES).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: WHEAT DP PROTEINS AND USES THEREOF

```
   1 GAATTCGGCACGAGCCGCAATGGCGCCTCCCCGCGGCGGAGCTGCTGCGGCCGCTACCGCC   61
   1                 M  A  P  P  R  G  G  A  A  A  A  A  T  A   14
  61 GCACTGGACCTGACCGGCGTGCACATTCTCGAAGCTTCCAGTGTCCCCCCGCTTCCCGAA 120
  15   A  L  D  L  T  G  V  H  I  L  E  A  S  S  V  P  P  L  P  E   34
 121 CGCGGCGGTAATGCGGTCCAAAGGAAGGGGGCTGTTGACCCGGATAAAGATAGGAAGAAG 180
  35   R  G  G  N  A  V  Q  R  K  G  A  V  D  P  D  K  D  R  K  K   54
 181 GAGAAGGCTGCGGCACCGAGGATCACCGGTTGGGGGCTCCGCGAGTACAGCAAAATAGTT 240
  55   E  K  A  A  A  P  R  I  T  G  W  G  L  R  E  Y  S  K  I  V   74
 241 TGTGAGAAAGTTGAAGCCAAAGGAAGAACAACATACAATGAGGTTGCAGACGAAATTTAT 300
  75   C  E  K  V  E  A  K  G  R  T  T  Y  N  E  V  A  D  E  I  Y   94
 301 TCAGAGCTGAAGTCCATGGCACATATTGGTCAAGGGTTTGATGAGAAGAATATTAGGCGG 360
  95   S  E  L  K  S  M  A  H  I  G  Q  G  F  D  E  K  N  I  R  R  114
 361 AGAGTGTATGATGCTTTCAACGTTCTCATTGCACTTCGTGTTATTGCAAAAGAAAAAAAG 420
 115   R  V  Y  D  A  F  N  V  L  I  A  L  R  V  I  A  K  E  K  K  134
 421 GAGATACGGTGGATGGGCCTTTCAAATTACAGATATGAAAAAATAAAGAAGCTTGAGGAA 480
 135   E  I  R  W  M  G  L  S  N  Y  R  Y  E  K  I  K  K  L  E  E  154
 481 GTTCGTAAAGAACTCGTCAACAAGATTAGGAACAAGAAGGCACTCCTCCAGGAAATCGAA 540
 155   V  R  K  E  L  V  N  K  I  R  N  K  K  A  L  L  Q  E  I  E  174
 541 AAACAGTTTGATGATCTCCAAAACATCAAGTTACGTAACCAAACACTGGAAAGCTCAGCA 600
 175   K  Q  F  D  D  L  Q  N  I  K  L  R  N  Q  T  L  E  S  S  A  194
 601 GAGAATGTTAATGGCATCCGCCTTCCATTCGTATTGGTCAAGACATCTAGGAAAGCAAGG 660
 195   E  N  V  N  G  I  R  L  P  F  V  L  V  K  T  S  R  K  A  R  214
 661 GTGGAAATTGAGATTTCAGATGACTCGAAGTTTGCCCATTTCGAGTTCAATGGTGCACCA 720
 215   V  E  I  E  I  S  D  D  S  K  F  A  H  F  E  F  N  G  A  P  234
 721 TTCACATTGCATGATGATCTCTCAATCCTTGAGGGGGTAAGGCGTAACAGCATAGGAAGA 780
 235   F  T  L  H  D  D  L  S  I  L  E  G  V  R  R  N  S  I  G  R  254
 781 GCTGGCCGCGCCACCCTTCACTAGAGACTCAAGAATATTACAAATGAATTAAAAGTGTTA 840
 255   A  G  R  A  T  L  H  *261
 841 GAACTGGCACAGCCGGATTCTTTTGCACAGCTATGTATAGCTATATATCCTCATGAAAAC 900
 901 TTGACCTAGTTTATAGGACAGTCTCTCAGGCTTGAGAAGATTTTAACCTGCAAATTTTGT 960
 961 CTCCTTTTTGTGCCTAGCAGGTTATTAGGTCTCAGATAGATGATTCATATATGTGCTGCT 1020
1021 ATGAAAACATTGATAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1080
1081 AAAAAAAA1089
```

Sequence of TmDP cDNA and deduced amino acid sequence.

(57) Abstract: A method of controlling plant cell cycle is provided characterised in that it comprises increasing or decreasing E2F-dimerization partner (DP) protein activity in a plant cell through expression of a recombinant DP peptide or protein in that cell. Further provided is use of such proteins in identifying genes involved in cell cycle control.

WO 01/21644 A2

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WHEAT DP PROTEINS AND USES THEREOF

The present invention relates to novel nucleic acids and transcription factor proteins and peptides encoded thereby which have properties of modifying plant cell cycle when expressed or otherwise incorporated into plant cells. Such modification can be used to manipulate plant and plant organ or tissue size or to overproduce specific gene products. Particularly provided are recombinant plant, and particularly wheat Triticum monococcum, DP protein transcription factor encoding nucleic acids, modified forms thereof and antisense nucleic acids thereto.

The present invention further provides methods for identifying and/or isolating DNA responsive to said transcription factor proteins, particularly identifying and/or isolating such DNA in the form of promoters with or without open reading frames (ORFs) and other regulatory regions, eg. as whole or partial genes. It particularly relates to methods of identifying and/or isolating such ORFs, promoters and whole genes responsive to E2F in hetero-dimer form with its dimerization partner protein (DP), and most particularly such DNA as found in plants. By identifying such DNAs in plants it is thus made possible using molecular biology and computational techniques to identify homologous DNAs of the same or similar function in other organisms, particularly animals, viruses and fungi, but also yeast and bacteria.

In animal cells, the retinoblastoma (Rb) protein is a major regulator of the cell cycle transition from G1 to S phase. Rb exerts its function by regulating the activity of the so-called E2F family of transcription factors which control the expression of a set of genes required for G1 passage and the G1/S transition. In human cells, six E2F proteins have been identified (E2F-1 to-6). Full activity of E2F members depends of heterodimerization with other less-related members, the DP (dimerization partner) proteins, of which two members (DP-1 and DP-2 or DP-3, depending on the nomenclature) have been identified.

In plant cells, the discovery, isolation and characterization of a maize Rb related protein (ZmRBR, formerly ZmRb1) revealed that a Rb-like pathway seems to be crucial for cell cycle transitions. Further studies strongly suggest that components

of this pathway might also be responsible for maintenance of particular differentiated states in plants.

WO 99/53075 (E.I. Du Pont de Nemours), incorporated herein by reference, discloses three Expressed Sequence Tags (ESTs) derived respectively from Impatiens balsamina developing seed, from etiolated Corn seedlings 14 days after planting and from Soybean 8 day old root inoculated with eggs of cyst nematode Heterodera glycines (Race 14) for 4 days (see Example 5 therein). These respective ESTs have relative low but apparently significant similarity to known Xenopus laevis (46%) and Mus musculus (37%) and (48%) respectively. Percentage similarity between the three plant ESTs was found to be between 31% and 78%. That patent teaches that BLAST score and probabilities indicate that these sequences represent the first (identified) plant sequences encoding (part of) DP-1 proteins. That patent application further discloses ESTs encoding parts of putative plant DP-2 proteins and E2Fs.

WO 00/47614 (Pioneer Hi-Bred), incorporated herein by reference, was published after the priority dates of the present application but before its filing date, and discloses a full cDNA and amino acid sequences for a Zea Mays DP protein, but does not designate this as of any of the sub-groups DP-1 to DP-3. The patent application provides a most comprehensive set of strategies and protocols for use of such cDNA and protein, but provides no description of them having actually been expressed or produced. No properties for the putative DP protein are given other than sequence information and thus it would appear that the protein amino acid sequence at least is merely that deduced from the nucelotide sequence of the DNA.

The present applicant's copending patent application WO 99/58681 relates to recombinant DNA derived from Triticum monococcum that encodes for a protein that acts as a plant E2F transcription factor. Although the determination of the nucleic acid sequence that will encode a functional plant E2F factor allows for transformation of plants such as to control cell cycle, and thus plant and specific plant organ and tissue size, it is thought to be possible that significantly high overexpression of E2F alone could be deleterious based on analogous animal

studies. These show that such expression might trigger an apoptotic-like pathway wherein plant cells would detect an increase in E2F and a concomitant entry in S-phase as abnormal.

In human systems binding of E2F to its binding sites is strongly stimulated (by over 50 fold) by presence of the DP protein together with human E2F in experimental systems. Binding of plant E2F to its responsive sites may be made more efficient and/or specific in the presence of such a plant DP. In human cells, a number of cell cycle and DNA replication-related genes have been shown to be expressed in an E2F-dependent manner. Thus the use of plant E2F as a trans-activator in plants would be more efficient when combined with DP, particularly a plant DP, in the same plant if such a partner were to be available.

As yet a systematic, genome-wide search for all the E2F-responsive genes has not been carried out. One of the main reasons is that any protocol designed to randomly isolate DNA fragments containing E2F-binding sites does not identify the promoter regions, where they are located and/or the genes that they regulate. Furthermore, the lack of whole genomic information is a drawback of such approaches. Attempts to identify potential E2F-binding sites have been carried out using the so-called CASTing (cyclic amplification and selection targets) system (Ouellette et al. (1992) Oncogene 7, 1075-1081;Tao et al., (1997) Mol. Cell Biol. 17, 6994-7007). This has allowed the identification of oligonucleotide sequences that contain E2F-binding sites but, again, no information on the promoters and where they are located can be extrapolated.

The Arabidopsis thaliana genome has now been largely made available, and is shortly to be fully sequenced, with for example an Arabdopsis sequence having similarity to Homo-sapiens DP-2 protein genes having been deposited on 3rd April 2000 at EMBL under ID code ATT22P11. The availability of cDNA clones encoding functionally interacting plant E2F and DP proteins from the same plant speices, ie. Triticum monococcum, provided now by the present inventors, offers a unique opportunity for a genome-wide search, identification and isolation of all E2F-binding

sites present in a plant genome. Such identified genes may then, for example, be used in BLAST searches of the Arabdopsis genome to establish existence of important, so far unidentified, universally applicable plant homologous gene families that may serve as targets for selection, molecular biological transformation and chemical agents.

The DP encoding DNA and protein encoded thereby provided by the present inventors are found to differ significantly in sequence from those of the prior art identified above. Thus, when compared using BLAST the present DP sequence has less than 45% identity and less than 60% similarity (homology) with the WO 99/53075 Impatiens DP1 EST where they overlap. When compared with the WO 99/53075 DP-2 EST from Zea Mays the identity found is less than 48% and the similarity is less than 58% where they overlap. When compared with the Triticum aestivum DP-2 EST identity found is less than 44% and similarity is less than 57%.

As described above, E2F and DP are two proteins that hetero-dimerize to form an active transcription factor that regulates the transition from G1 to S phase of the cell cycle and, later, the expression of genes required for S-phase progression. It is further known that E2F and retinoblastoma (Rb) proteins interact as a hetero-dimer in cells to repress certain genes. This repression involves binding of the retinoblastoma protein to the E2F-DP hetero-dimer that is in turn bound to sites on DNA through the E2F DNA binding domain. Thus certain important G1 repressed genes are thought to be identifiable best by a combination of E2F and Rb or Rb-E2F-DP together.

The determination of a functional plant DP amino acid sequence and the nucleic acid sequences encoding for this provides the possibility for transforming plants whereby E2F and DP levels or activity may be co-ordinated, thus avoiding any deleterious effects and allowing for increased options for plant growth regulation. Particularly this determination allows for the co-overexpression, co-underexpression or opposite sense expression of proteins or peptides having plant E2F and DP activity.

The uses of such recombinant nucleic acids are potentially numerous. For

example, some plant organs that need a period of proliferation to increase cell number before differentiation, e.g. such as buds for flowers, shoots and leaves, could be increased in size if expression of E2F and DP is co-ordinated and the proliferation period is extended for a few more cycles. Use of temporally controllable promoters or naturally occuring modulators of expression of the recombinant nucleic acid would then allow the recombinant genes to be switched off and let differentiation occur. The descriptions of WO 99/53075 and WO 00/47614, incorporated herein by reference, provide further detailed instruction for uses for plant DP protein and encoding DNA that have equal application to the present materials.

The present inventors have now isolated, cloned and characterized a nucleic acid comprising a wheat (Triticum monococcum-Tm) derived cDNA encoding a plant protein which interacts with plant E2F from wheat (Triticum monococcum-Tm) in the yeast two-hybrid system. They have established that this cDNA clone encodes a plant E2F dimerization partner (DP) family member (TmDP) with amino acid regions having homology to conserved parts of animal DP proteins.

The present inventors have further provided a method for identifying and/or isolating DNAs corresponding to complete or partial genes that are regulated in G1 passsage, G1/ S-phase transition and/or S phase progression of the cell cycle, said method comprising contacting a sample of DNA, particularly whole genomic DNA that has been fragmented, eg by digestion or shearing, with a binding material specific for binding such complete or partial genes, removing non-bound DNA from the specific binding material then, releasing and isolating the bound DNA characterised in that the specific binding material comprises a peptide or protein including the DNA binding sequence of a protein that is capable of acting as a part of a plant hetero-oligomer transcription activator or repressor, particularly as part of the wheat E2F/ DP heterodimer and most particularly is wheat Tm E2F/DP heterodimer.

With respect to the present specification and claims, the following technical terms are used in accordance with the definitions below.

A "functional variant" of a peptide or protein is a polypeptide the amino acid sequence of which can be derived from the amino acid sequence of the original peptide or protein by the substitution, deletion and/or addition of one or more amino acid residue in a way that, in spite of the change in the amino acid sequence, the functional variant retains at least a part of at least one of the biological activities of the original protein that is detectable for a person skilled in the art. A functional variant is generally at least 50% homologous, advantageously at least 70% homologous and even more advantageously at least 90% homologous to the protein from which it can be derived. Preferably the amino acid sequence of the functional variant is 50% identical, more preferably 70% identical and most preferably 90% identical to the peptide or protein. Any functional part of a protein or a variant thereof is also termed functional variant. Further preferred conservatively substituted variants are as defined in WO 00/47614, pages 7-8. Biological function in the present application is the ability to bind to and or alter the DNA binding capabilities of E2F proteins.

By homologous is meant that the stated percentage of the amino acid sequence has identity or is of conservatively substituted amino acids. By identical is meant that the stated percentage of the amino acid sequence is identical. Both these percentage terms allow for gapping of sequences to allow alignment as is described below.

Particularly, by the term identity is meant that the stated percentage of the claimed amino acid sequence or base sequence is to be found in the reference sequence in the same relative positions when the sequences are optimally aligned, notwithstanding the fact that the sequences may have deletions or additions in certain positions requiring introduction of gaps to allow alignment of the highest percentage of amino acids or bases. Preferably the sequence are aligned by using 20 or less gaps, ie. the total number of gaps introduced into the two sequences when added together is 20 or less, more preferably 10 or less. The length of such gaps is not of particular importance as long as one or other of the two defined activities relating to E2F binding and E2F DNA binding modulation is retained but generally will be no more than 50, and preferably no more than 10 amino acids, or 150 and preferably no more than 30 bases.

Variants from the aforesaid sequences preferably are conservative substitutions. The expression 'conservative substitutions' as used with respect to amino acids relates to the substitution of a given amino acid by an amino acid having physicochemical characteristics in the same class. Thus where an amino acid has a hydrophobic characterising group, a conservative substitution replaces it by another amino acid also having a hydrophobic characterising group; other such classes are those where the characterising group is hydrophilic, cationic, anionic or contains a thiol or thioether. Such substitutions are only contemplated where the resultant protein has activity as a DP peptide or protein as discussed with respect to E2F heterodimerization and modulation of E2F-DNA binding or transcription activation.

Algorithms and software suitable for use in aligning amino acid or nucleotide sequences for comparison and calculation of sequence homology or identity will be known to those skilled in the art. Significant examples of such tools are the Pearson and Lipman search based FAST and BLAST programs. Details of these may be found in Altschul et al (1990), J. Mol. Biol. 215: 403-10; Lipman D J and Pearson W R (1985) Science 227, p1435-41. Publically available details of BLAST may be found on the internet at 'http://www.ncbi. nlm.nih.gov/BLAST/blast-help.html'. Thus such homology and identity percentages can be ascertained using commercially or publically available software packages incorporating, for example, FASTA and BLASTn software or by computer servers on the internet. Examples of the former are the GCG program package (Devereux et al Nucelic Acids Research (1984) 12 (1): 387) and the Bestfit program (Wisconsin Sequence Analysis Package, eg. Version 8 for Unix or IBM equivalent, Genetics Computer Group, University Researh Park, 575 Science Drive, Madison, WI 53711) which uses the local homology algorithm of Smith and Waterman, Advances in Mathematics 2:482-489 (1981). Many international institutes, eg. Genbank (see http://www.ncbi.nlm.nih.gov/BLAST) and EMBL: (see http://www.embl-heidelberg.de/Blast2), offer internet services.

Parameters used in with software packages and internet servers should be applied with the appropriate sequence lengths and aforesaid gap characteristics in

mind. Alignment strategies are discussed further in WO 98/40483 on pages 39 to 41, which document is incorporated herein by reference

Convenient parameters for BLAST searches are the default values, ie. for EMBL Advanced Blast2: Blastp Matrix BLOSUMS 62, Filter default, Echofilter X, Expect 10, Cutoff default, Strand both, Descriptions 50, Alignments 50. For BLASTn defaults are again preferably used. GCG Wisconsin Package defaults are Gap Weight 12, Length weight 4. FASTDB parameters used for a further preferred method of homology calculation are mismatch penalty = 1.00, gap penalty =1.00, gap size penalty = 0.33 and joining penalty =30.0.

The term "overproducing" is used herein in the most general sense possible. A special type of molecule, usually a polypeptide or an RNA, is said to be "overproduced" in a cell if it is produced at a level significantly and detectably higher (e.g. 20% higher) than natural level, ie. that found in a cell of the same lineage that has not been transformed with plant the DNA described herein. Overproduction of a molecule in a cell can be achieved via both traditional mutation and selection techniques and genetic manipulation methods as long as one of the parents used in selection methods is of transgenic type. The term 'underproducing' is intended to cover production of polypeptide or mRNA at a level significantly lower than the natural level (eg. 20% or more lower), but particularly to undetectable levels.

The term "ectopic expression" is used herein to designate a special realisation of overproduction in the sense that, for example, an ectopically expressed peptide or protein is produced at a spatial point of a plant where it is naturally not at all (or not detectably) expressed, that is, said peptide or protein is overproduced at said point. Particularly preferred ectopic expression is that which only reaches functional levels in a selected tissue and does not do so throughout the plant. This preferred ectopic expression is in contrast to constitutive expression.

A "gene" is a DNA sequence that controls a discrete hereditary characteristic and as such is not limited to DNA coding for proteins but relates to ORFs together with any regulatory sequence thereof. Thus a partial gene or part of a gene may include no ORF sequence, but consist only of regulatory, eg. a promoter, or associated

sequence, eg introns, particularly being in the present case that sequence which is subject to transcription factor specific binding, particularly by E2F-DP hetero-dimer and Rb-E2F-DP tertiary complex.

The expression 'conservatively substituted' as used with respect to amino acids relates to the substitution of a given amino acid by an amino acid having physicochemical characteristics in the same class. Thus where an amino acid has a hydrophobic characterising group, a conservative substitution replaces it by another amino acid also having a hydrophobic characterising group; other such classes are those where the characterising group is hydrophilic, cationic, anionic or contains a thiol or thioether. Such substitutions are only contemplated where the resultant protein has activity as a DP peptide or protein as discussed with respect to DNA and E2F dimerization.

Nucleic acids of the invention may be degeneratively substituted with respect to that exemplified herein in the sequence listing. The expression 'degeneratively substituted' refers to substitutions of nucleotides by those which result in codons encoding for the same amino acid; such degenerative substitutions being advantageous where the cell or vector expressing the protein is of such different type to the DNA source organism cell that it has different codon preferences for transcription/translation to that of the cDNA source cell. Such degenerative substitutions will thus be host specific.

Disclosure of the Invention

In a first aspect of the present invention there is provided a method of controlling one or more of plant growth, gene expression, cellular DNA replication, cell cycle progression, differentiation and development comprising increasing or decreasing E2F-dimerization partner (DP) protein activity in a plant cell through expression of a recombinant DP peptide or protein in that cell characterised in that the peptide or protein comprises a sequence SEQ ID No 2, a functional part thereof, or a sequence having at least 70% homology to either, that peptide or protein being

capable of interacting with a plant E2F protein or peptide such as to alter E2F activity in the plant cell.

Preferably the peptide or protein is of 50% or more identity with that of the corresponding full length or part of SEQ ID No 2, more preferably 70% identity and most preferably 90% identity. More preferably the peptide or protein is of 90% homology with the full length or part of SEQ ID No 2.

Preferably the method is characterised in that the plant DP activity comprises one or both of (i) the ability to dimerize with plant E2F protein and (ii) the ability to modulate, particularly enhance, E2F binding to E2F/DP transcription factor binding sites in plant DNA.

The method may include steps of altering the plant DP protein level, the E2F-DP DNA-binding activity, transactivation properties, and/or the DP/E2F-binding activity. The plant DP may be modified alone and/or in combination with a modification of the levels or activity of plant E2F and/or plant Rb. The ability to enhance E2F binding to the E2F transcription factor binding sites in plant DNA need not necessarily lead to transcription activation. Binding of the E2F-DP heterodimer with inhibition of such activation can also be provided using the present invention as can be decreased E2F-DNA binding or tracnscription.

Particularly the method may be used to alter plant cell, organ or tissue shape, and it may particularly alter cell proliferation characteristics such as to increase or decrease plant cell, organ or tissue size. The method may also increase or decrease expression of other proteins with transformed cells and cells derived therefrom, particularly direct or indirect progeny.

In a second aspect the present invention provides an isolated, enriched, cell free and/or recombinantly produced protein or peptide, capable of altering E2F-dimerization partner (DP) activity in a plant cell, characterised in that it has one or both DP activities in plants selected from (i) the ability to dimerize with plant E2F protein and (ii) the ability to modulate, particularly enhance, E2F binding to E2F transcription factor binding sites in plant DNA or effect thereof.

characterised in that the protein or peptide comprises an amino acid as shown in SEQ ID No 2 or a functionally active part thereof or a sequence having at least 70% homology to such sequence or part, more preferably at least 90% homology thereto and most preferably having at least 50% identity therewith, or still more preferably having at least 70% or at least 90% identity therewith.

More preferably the peptide or protein comprises at least 50% of the contiguous sequence and still more preferably at least 70% thereof.

One group of possible peptides or proteins of the invention are characterised in that they are of SEQ ID No 2 or variants thereof modified such that the amino acid sequence is mutated such that its ability to dimerize with E2F protein is reduced from that of the native sequence or abolished completely therefrom, whereby the peptide is capable of acting as a DP protein which decreases or abolishes native or recombinant E2F binding to its DNA binding site, thus inhibiting or abolishing E2F activity in a cell in which is its present.

Preferred peptides or proteins of the invention are further characterised in that they comprises a sequence found in that of SEQ ID No 2 or having at least 70% homology thereto selected from those comprising

SEQ ID No 6 ARAAMAPPRGGAAAAATAALDLTGVHILEAS SVPPLPE RGGNAVQRKGAVDP

SEQ ID No 8 DKDRKKEKAAAPRITGWGLREYSKIVCEKVEAKGRT TY NEVADEIYSELKS

SEQ ID No 10 MAHIGQGFDEKNIRRRVYDAFNVLIALRVIAKEKKEIR W MGLSNYRYEKIKKLEEV

SEQ ID No 12 RKELVNKIRNKKALLQEIEKQFDDLQNIKLRNQTLESS A ENVNGIRLPFVLVKTSR

SEQ ID No 14 KARVEIEISDDSKFAHFEFNGAPFTLHDDLSILEGVRGNS IGKAGRATLH

Most preferably the sequence comprises two or more of these sequences or sequences at least 70% homologous, more preferably at least 90% and still more

preferably at least 95% homologous thereto, still more preferably being the stated at least percentages identical.

Particularly useful peptides or proteins comprise at least one or more of the sequences of SEQ ID No 2 or percentage homologous or percentage identical sequences thereto comprising amino acids 70 to 136 (the so called DNA binding domain), amino acids 137 to 200, (the heterodimerization domain) and amino acids 55 to 62 (the putative nuclear localization signal), more preferably two or more of these.

For some purposes it will be convenient to provide peptides or proteins of reduced length, for example 16 to 300, more preferably from 16 to 100 amino acids.

Useful variants of such proteins however are those in which non-essential or essential amino acids for E2F dimerization are modified, eg. by site directed mutagenesis, eg using PCR.

Particularly useful is nucleic acid the expression of which is controlled using tissue specific or chemically inducible promoters

A third aspect of the present invention provides isolated, enriched, cell free and/or recombinant nucleic acid comprising a sequence encoding for expression of a protein or peptide as described in the first aspect of the invention. Preferred nucleic acids comprise DNA of less than 4,000 basepairs. Preferred nucleic acids comprise only one peptide or protein encoding DNA sequence, optionally together with a reporter gene.

Preferably the nucleic acid is that encoding for a plant DP or a functional variant thereof including the coding nucleic acid sequence of SEQ ID No 1 or a part thereof encoding for all or a functional part of the amino acid sequence shown therein as defined above. Preferred nucleic acid comprises DNA or RNA wherein when the nucleic acid is RNA the base T is substituted by U.

A nucleic acid encoding for a TmDP of sequence of SEQ ID No 1 has been deposited on August 17th 1999 under the terms of the Budapest Treaty for the International Recognition of Microorganism Deposits for Patent Purposes of 28th April 1977 at the Coleccion Espanola de Cultivos Tipo in plasmid pCLON33 under

deposit number CECT 5195. Restriction enzymes contained in the multicloning site of the plasmid can be used to excise the insert cDNA from this, eg. EcoRI and XhoI or BamHI and XhoI. For in vitro transcription-translation, the full-length TmDP DNA was cloned into pBluescriptSK+ using BamHI and XhoI restriction enzymes.

It will be understood that nucleic acids of the invention may be double stranded DNAs or single stranded DNA of the cDNA or a sequence complementary thereto, eg. such as will have use as a probe or primer.

Preferred nucleic acids are characterised in that they encode for a plant DP or a functional variant thereof including the sequence of SEQ ID No 1 or a sequence complementary thereto. Further preferred nucleic acids comprise DNA, whether double or single stranded, sense, complementary or otherwise antisense thereto. Preferred nucleic acids comprise a cDNA optionally provided together with promoter, enhancer or stop sequences with no other gene coding regions. .

The DNA or RNA of the invention may have a sequence containing degenerate substitutions in the nucleotides of the codons in the sequences encoding for DP proteins or peptides of the invention. In RNA U's replace the T's of DNA. Preferred *per se* DNAs or RNAs are capable of hybridising with the polynucleotides encoding for peptides or proteins of the invention in conditions of low stringency, being preferably also capable of such hybridisation in conditions of high stringency.

The terms "conditions of low stringency" and "conditions of high stringency" are of course understood fully by those skilled in the art, but are conveniently exemplified in US 5202257, columns 9 and 10 and in WO 98/40483 on page 3; both of which are incorporated herein by reference. Thus, generally, the most preferred nucleic acids of the invention will hybridise at the most stringent conditions described in these patents while other embodiments will hybridise at the milder stringency or low stringency conditions. Further examples of preferred stringency are described in PCT/IB97/00409, see page 21, line 23 to page 27, line 15 and the corresponding US patent application, incorporated herein by reference.

In US 5202257 low-stringency conditions comprise a temperature of about

37°C or less, a formamide concentration of less than about 50%, and a moderate to low salt (SSC) concentration; or, alternatively, a temperature of about 50°C or less, and a moderate to high salt (SSPE) concentration. The preferred conditions for such screening comprise a temperature of about 37°C, a formamide concentration of about 20%, and a salt concentration of about 5 times standard saline citrate (SSC; 20 times. SSC contains 3M sodium chloride, 0.3M sodium citrate, pH 7.0); or a temperature of about 50°C, and a salt concentration of about 2 times SSPE (1 times SSPE contains 180 mM NaCl, 9 mM $Na_2 HPO_4$, 0.9 mM NaH_2PO_4 and 1 mM EDTA, pH 7.4.

High stringency conditions are described as comprising a temperature of about 42°C or less, a formamide concentration of less than about 20%, and a low salt (SSC) concentration; or, alternatively, a temperature of about 65°C or less, and a low salt (SSPE) concentration. The preferred conditions for such screening are described as comprising a temperature of about 42°C, a formamide concentration of about 20%, and a salt concentration of about 2.times. SSC; or a temperature of about 65°C, and a salt concentration of about 0.2 times SSPE.

WO 00/47614 also describes conditions of stringent hybridization conditions of high, moderate and low nature and these are found on its page 17-18, incorporated herein by reference. These are thus further conventional equivalents for use in hybridaizations and are incorporated herein by reference for the purpose pf providing alternative option for identifying suiatble sequences.

Where modifications are made they should lead to the expression of a protein with different amino acids in the same class as the corresponding amino acids to these DP peptide or protein sequences; that is to say, they are conservative substitutions. Such substitutions are known to those skilled in the art see, for example, US 5380712 which is incorporated herein by reference, and are considered only when the protein is active as a DP peptide or protein with regard to its interactions with E2F as an E2F-DP heterodimer.

DNA or RNA provided from a plant or the deposit referred to above may be altered by mutagenic means such as the use of mutagenic polymerase chain reaction

primers. Methods of producing the proteins or peptides of the invention characterised in that they comprise use of the DNA or RNA of the invention to express them from cells are also provided in this aspect. Examples of probes are the DNA sequences corresponding to amino acid sequences SEQ ID No 2 to 8 above.

For the purpose of screening for plant DPs, a process which has heretofor been hampered due to human E2F and DP dissimilarity to plant E2F and DP, nucleic acid probes or primers comprising a double or single stranded DNA of sequence corresponding to 10 or more contiguous nucleotides taken from the sequence SEQ ID No 1 are provided, with the proviso that they are not selected from those just encoding for the amino acid sequence that is relatively highly conserved with human DP, ie.the DNA binding region of amino acid 70 to 136 is the most conserved region. Such probes and primers may be used in Northern and Southern blotting and in PCR, including RT-PCR, and LCR.

Oligonucleotides for use as probes conveniently comprise at least 18 contiguous bases of the sequences of the invention, preferably being of 30 to 100 bases long, but may be of any length up to the complete sequence or even longer. For use as PCR or LCR primers the oligonucleotide preferably is of 10 to 20 bases long but may be longer. Primers should be single stranded but probes may be also be double stranded ie. including complementary sequences.

For the purpose of downregulating native plant DP expression there is also provided antisense DNA to any of the nucleic acids of the invention described above. This technique is well known in the art but is generally illustrated by US 5356799 and US 5107065 by way of example, each of which is incorporated herein by reference. Anti-sense DNA is of length sufficiently long enough such that when expressed as RNA that downregulates expression of native or recombinant DP to levels that are measurably lower, eg. such that DP mRNA in Northern blots is lower or not measurably present.

A preferred form of the nucleic acid of the invention provides the DP protein or peptide encoding sequence as described above together with a sequence encoding

the E2F protein or peptide. Such sequences are conveniently and sometimes advantageously under control of the same regulatory element or elements, eg. promoters, such that they may act together as a heterodimer.

A fourth aspect of the invention provides a nucleic acid vector or construct comprising a nucleic acid of the present invention or comprising antisense nucleic acid thereto. Suitable vectors or constructs for introducing the peptides or proteins of the invention into plants will occur to those skilled in the art of plant molecular biology, but are conveniently those discussed below with respect to methods for producing transgenic plants. Such vector or construct may thus also comprise both DP and E2F peptides or proteins or antisense or other related sequences as described above.

A fifth aspect of the present invention provides a plant cell comprising recombinant nucleic acid, preferably recombinant DNA, of the third aspect of the invention. Nucleic acids of the invention are particularly provided in the form of such nucleic acid vectors or DNA construct comprising that nucleic acid or antisense nucleic acid sequence thereto.

A sixth aspect of the present invention provides a plant cell comprising antisense nucleic acid thereto capable of downregulating expression of native plant DP.

A seventh aspect of the present invention comprises a transgenic plant or part thereof comprising recombinant nucleic acid, a vector, DNA construct or cell as described above.

It will be realised that a most effective method of delivering proteins and peptides of the invention to plant cells is by expressing nucleic acid encoding them *in situ*. Such method is conventionally carried out by incorporating oligonucleotides or polynucleotides, having sequences encoding the peptide or protein, into the plant cell DNA. Such nucleotides can also be used to downregulate native DP expression by gene silencing coexpression or through antisense strategy. By use of mutagenesis techniques, eg. such as SDM, the nucleotides of the invention may be designed and

produced to encode proteins and peptides which are functional variants or otherwise overactivated or inactivated, eg. with respect to binding, of the invention

Preferred plants of the seventh aspect may comprise the nucleic acid of the invention in a construct in functional association with promoter, activating or otherwise regulating sequences. Preferred promoters may be tissue specific such that the resultant expression of peptide, and thus its effects, are localised to a desired tisssue. Promoters with a degree of tissue specificity will be known to those skilled in the art of plant molecular biology. Some of these are discussed below.

Methods of producing isolated or purified DNA/RNA, vectors and constructs capable of being used in the present invention will occur to those skilled in the art in the light of conventional molecular biology techniques. DNA, RNA and vector containing or encoding for these may be introduced into target cells in known fashion in the field of plant cell transformation. For example the method of introducing the DNA or RNA into cells, which eg. may be somatic or pollen cells, using techniques such as electroporation or gene gun technology.

It may be preferred to express the DNA or RNA of the invention throughout the plant, but in the event that tissue specific effect is to be exploited then it will be understood by those skilled in the art that tissue specific promoters, enhancers or other activators should be incorporated into the transgenic cells employed in operative relation with the DNA.

It will be realised by those skilled in the art that suitable promotors may be active ectopically, continuously or may be inducible. It will be appreciated by those skilled in the art that inducible or tissue specific ie promotors will have advantage in so far as they are capable of providing alteration of the aforesaid DP peptide or protein activity only when or where required, eg. at a particular stage of cell development or in a tissue such as leaves, roots, fruit or seeds or subparts thereof, eg. endosperm, that may be the subject of desired increase or decrease in size or even deletion.

No particular limitation on the type of promoter to be employed is envisioned,

although a reasonable amount of experimental trial may be expected to be undertaken to produce good results. Examples of tissue specific and inducible promoters can be found in the following patent literature: US 5086169 (pollen specific), US 5459252 and US 5633363 (root specific), US 5097025 ((i)seed, (ii)mature plant), US 5589610 (stamen), US 5428146 (wound), US 5391725 ((i)chloroplast, (ii) cytosol), US 4886753 (root nodule), US 4710461 (pollen), US 5670349 (pathogen), US 5646333 (epidermis), US 5110732 ((i) root , (ii) radical), US 5859328 (pistil), US 5187267 (heat shock), US 5618988 (storage organ), US 5401836 and US 5792925 (root), US 4943674 (fruit), US 5689044 and US 5654414 (chemical), US 5495007 (phloem), US 5589583 (meristem), US 5824857 (vasculature), each of which is incorporated herein by reference. Constitutive promoters will be well known to those skilled in the art and are discussed in the documents above and referred to below but for example include CaMv35S and alfalfa (MsH3g1) (see WO 97/20058 incorporated herein by reference).

Numerous specific examples of methods used to produce transgenic plants by the insertion of cDNA in conjunction with suitable regulatory sequences will be known to those skilled in the art. Plant transformation vectors have been described by Denecke et al (1992) EMBO J. 11, 2345-2355 and their further use to produce transgenic plants producing trehalose described in US Patent Application Serial No. 08/290,301. EP 0339009 B1 and US 5250515 describe strategies for inserting heterologous genes into plants (see columns 8 to 26 of US 5250515). Electroporation of pollen to produce both transgenic monocotyledonous and dicotyledonous plants is described in US 5629183, US 7530485 and US 7350356. Further details may be found in reference works such as Recombinant Gene Expression Protocols. (1997) Edit Rocky S. Tuan. Humana Press. ISBN 0-89603-333-3; 0-89603-480-1. All of these documents are incorporated herein by reference It will be realised that no particular limitation on the type of transgenic plant to be provided is envisaged; all classes of plant, monocot or dicot, may be produced in transgenic form incorporating the nucleic acid of the invention such that DP activity in the plant is altered, constituitively or ectopically.

In an eighth aspect of the present invention the present inventors make available antibodies capable of specifically binding with plant DP factor peptides or proteins of the first aspect of the present invention, thus enabling the identification and isolation of further peptides and proteins of the invention and nucleic acid sequences encoding therefor, eg. using techniques such as Western blotting. Preferably these antibodies are selected such that they do not bind to the other DPs described in Figure 2, eg. by selection for absence of significant binding to a column on which these are mounted as ligand binding agents. Such antibodies are provided by use of oligopeptides and polypeptides consisting of parts of the TmDP that are not replicated in mammalian DPs, ie. not the so called DNA binding domain specified above, to raise the antibody eg in rats or rabbits etc..

In a ninth aspect of the invention there is provided the aforesaid method for identifying and/or isolating DNAs corresponding to complete or partial genes that are regulated in G1 passsage, G1/S-phase transition and/or S phase progression of the cell cycle, said method comprising contacting a sample of DNA, particularly whole genomic DNA that has been fragmented, eg by digestion or shearing, with a binding material specific for binding such complete or partial genes, removing non-bound DNA from the specific binding material then, releasing and isolating the bound DNA.

characterised in that the specific binding material comprises a peptide or protein including the DNA binding sequence of a protein that is capable of acting as a part of a plant hetero-oligomer transcription activator or repressor.

Preferably the method is characterised in that the specific binding material comprises a peptide or protein which includes a plant E2F DNA binding domain, however, other plant hetero-oligomer transcription activator protein DNA binding sequences may be used, eg, the GRAB proteins of PCT/EP98/03662 incorporated herein by reference.

The specific binding material particularly preferably comprises a peptide or protein which includes a plant E2F DNA binding domain together with a plant E2F-dimerization partner (DP) hetero-dimerization domain and/or a plant retinoblastoma protein E2F binding domain.

Most advantageously the material comprises a peptide or protein that comprises a plant E2F DNA binding domain together with a plant E2F-dimerization partner (DP) binding domain, hereinafter called the E2F hetero-dimerization domain, optionally together with a peptide or protein that includes a plant dimerization partner (DP) E2F binding domain, hereinafter called the DP hetero-dimerization domain.

Preferably these are both the TmDP and TmE2F or functional variants or parts thereof as defined above having the set defined homology and/or identity.

Preferably the specific binding material peptide or protein is labelled or tagged to assist in identifying or immobilising it, particularly when in bound complex with the gene to be identified and/or isolated. More preferably, and particularly advantageously, the specific binding material comprises two peptides or proteins, one including the E2F DNA binding domain, and one including the DP hetero-dimerization domain bound together as a hetero-dimer.

The peptide or protein including the E2F DNA binding domain may consist of a complete plant E2F protein, optionally but preferably labelled, but may consist of only a part thereof, eg. just the DNA binding domain, the DNA binding domain and the hetero-dimerization domain or larger peptides or proteins including these, eg. truncates of plant E2F. Suitable peptides and proteins are described in copending PCT/EP99/03158 (incorporated herein by reference) and include all the functional variants including the E2F binding domain disclosed therein. Preferably the peptide or protein is a truncated or whole wheat E2F protein as disclosed in that patent and having all or part of SEQ ID No 4 shown below, preferably conjugated with a label.

The peptide or protein including the DP hetero-dimerization domain may consist of a complete plant DP protein, optionally but preferably labelled, but may consist of only a part thereof, eg. just the hetero-dimerization-domain or larger peptides and proteins including this, eg. truncates of plant DP. Suitable peptides and proteins include functional variants. Preferably the peptide or protein is a truncated or whole wheat DP protein, having all or part of the amino acid sequence SEQ ID No 2 shown below, or a functional variant thereof, preferably conjugated with a label.

The binding material may alternatively or additionally incorporate a peptide or protein including a plant retinoblastoma protein E2F binding domain, hereinafter called a plant retinoblastoma (Rb) hetero-dimerization domain. Again, this may take the form of a whole plant retinoblastoma protein, optionally but preferably labelled, but may be a truncate thereof, or a functional variant of one of these.

For the purpose of merely identifying genes and parts thereof which bind with E2F on its own, a peptide or protein amino acid sequence comprising only the DNA binding domain of E2F may be present, optionally in a peptide or protein including other non-functional sequence. However, most advantageously two peptides or proteins will be present and will include the respective E2F and DP hetero-dimerization domains with the E2F DNA binding domain in order that the full enhancing effect of DP on E2F binding may be provided. It may be foreseen that use of a peptide or protein including the E2F DNA binding and hetero-dimerization domains together with a peptide or protein including the DP hetero-dimerization domain without the DNA binding enhancing domain may be of interest for some genes, it may be preferred to include also the DP DNA binding enhancing domain.

Labelling of the respective peptide or proteins may in principle be made with conventional labelling material, but advantageously these should be different for each of the peptides or proteins used. Conveniently the label is in the form of a fusion partner, eg, the peptide or proteins are provided as GST or MBP fusion peptide or proteins. Other forms of labelling will occur to those skilled in the art in the light of these. Use of fusions such as GST and MBP 'tags' allows for immobilisation of the peptide or protein on Gltathione Sepharose and on Amylose resins respectively.

Using the DNAs isolated by the present method it is possible to obtain their sequences and, by using computational homology analysis of other organism's genome eg of plant, animal, yeast, bacteria, virus or fungal genomes, or molecular biology probing or other analysis techniques to identify therapeutic or otherwise industrially useful target genes involved in cell cycle that are no readily obtainable due to the comparative complexity or lack of information of/on such systems.

The present inventors have cloned, over-expressed in E. coli and purified both E2F (wheat E2F) and DP (wheat, E2F) as GST and MBP fusion proteins, respectively. They have shown these to hetero-dimerize *in vitro* and the corresponding tags do not seem to interfere with this interaction. Furthermore, binding of E2F to a double-stranded oligonucleotide containing the consensus sequence for human E2F-1 is stimulated ~20-50-fold by its association with DP in hetero-dimer form.

A preferred method of the invention comprises a procedure for isolating DNA fragments containing E2F-binding sites wherein a column of the specific binding material, eg. the purified labelled E2F-DP hetero-dimer, is prepared and a solution holding a DNA sample to be screened for genes and gene parts is passed down it. Conveniently a GST-E2F/MBP-DP hetero-dimer is prepared, is used to coat a support material such as glass beads in order to maximise surface area, and used to load a column. The DNA sample is advantageously a digested purifed genomic DNA from a plant, eg. Wheat or Arabidopsis, the digestion having been carried out with a frequent cutter providing convenient DNA ends, e.g. a Sau3A cutter. Alternatively, genomic DNA can be sheared to a size ranging from 200 to 500 base pairs.

In a further preferred method the DNA sample is bound batchwise to the labelled hetero-dimer binding material, eg. on a support material or in free form, suspended in a buffer used routinely for binding experiments.

Washing of the column or batch of binding material is carried out extensively to remove non-specific binding DNAs before the specifically bound DNA is eluted with a high salt solution. If this removal is not complete, elution of the protein-DNA complexes may be carried out and the DNA may be purified in a separate step.

A library may be made of all bound DNA fragments and the inserts sequenced. Using conventional bioinformatic tools, the chromosomal location of each DNA fragment may be determined. The corresponding analysis of E2F/DP-dependent promoter activity may be determined using convenient reporter genes.

This procedure provides the capability to produce a virtually complete list of all the E2F, E2F/DP, Rb/E2F and/or Rb/E2F/DP responsive or repressed genes in the

genome of a multi-cellular organism, a plant in particular, something that as yet has been unachievable. In addition to the genes known to respond or bind to E2F/DP or Rb/E2F or Rb/E2F/DP in human cells, the present method can identify new or known genes in plants whose promoters are not suspected to be dependent on E2F and this information can also be used to identify counterparts in human cells.

Particularly the present method is characterised in that the plant DP activity comprises one or more of (i) the ability to dimerize with plant E2F protein and (ii) the ability to enhance or decrease E2F binding to E2F/DP transcription factor binding sites in plant DNA. The ability to enhance E2F binding to the E2F transcription factor binding sites in plant DNA need not necessarily lead to transcription activation.

Particularly the genes, promoters and ORFs provided by the present method may be used in recombinant form to alter cell, organ or tissue shape, particularly in plants but also in other organisms, and it may particularly alter cell proliferation characteristics such as to increase or decrease plant cell, organ or tissue size. The method may also increase or decrease expression of other proteins with transformed cells and cells derived therefrom, particularly direct or indirect progeny.

In a tenth aspect the present invention provides a specific binding material characterised in that it comprises

(i) a peptide or protein having DNA binding activity with respect to plant DNA transcription activator or repressor factor binding sites, particularly in genomic DNA, and having the ability to dimerize or oligomerize with a further such plant protein together with one or more of said further a peptides or proteins.

Preferably the specific binding material is characterised in that it comprises

(i) a peptide or protein having DNA binding activity with respect to plant DNA E2F transcription factor binding sites, particularly those present in genomic DNA and having the ability to dimerize with plant DP protein together with one or both of

(ii) a peptide or protein that is capable of binding to plant E2F through its DP hetero-dimerization domain and

(iii) a peptide or protein that is capable of binding to plant E2F through its retinoblastoma protein binding domain.

Preferably these are the wheat peptide or proteins and functional variants and parts thereof as define above.

Preferably the material comprises a hetero-dimer of (i) and (ii) or (i) and (iii). Preferably the binding material comprises both of the peptides or proteins (i) and (ii), most preferably as a hetero-dimer. The material may also comprise the peptides or proteins (i) and (ii) in the form of a hetero-dimer together with peptide or protein (iii). Most preferably peptides or proteins (i), (ii) and/or (iii) are in labelled or tagged form, particulary preferably being labelled through being fused with a tag eg. a fusion peptide or protein, advantageously independently, for example with GST or MBP.

In a preferred form of this aspect the binding material is provided in the form of a coating or otherwise bound form on a support material, eg. on glass beads, phosphocellulose, sepharose or amylose or some similar support such as is used on an affinity column, or particularly being in the form of an elutable column filled with the peptide or protein (i) and (ii) and/or (iii), particularly as said independently double labelled hetero-dimers comprising said fusion peptides or proteins and/or as a coating on support material. Alternatively the material is in the form of particles, eg. granules or spheroids, of the peptides or proteins and/or hetero-dimers described above.

As stated previously the DP, E2F and Rb peptides or proteins may be truncates and/or variants of the respective parts of the plant DP and E2F proteins set out in SEQ ID No 2 and SEQ ID No 4 herein below or the known plant Rb proteins in the prior art (see eg, sequences in copending WO/EP97/03070 incorporated herein by reference). Such variants preferably comprise respective parts of these amino acid sequences or a sequence having at least 50% identity therewith, or still more preferably having at least 70 or at least 90% identity and most preferably at least 95% identity therewith.

More preferably the peptide or protein comprises at least 50% of the contiguous sequence and still more preferably at least 70% thereof.

For the avoidance of doubt, particularly useful peptides or proteins for the DP function of the present invention comprise amino acids 137 to 200 of DP SEQ ID No 2, (the hetero-dimerization domain) and conservatively substituted variants thereof.

The peptides and proteins for use in the present method may be produced using the DNA described in the above described copending patents, particularly as expressed in whole or truncated form or as fusion with GST or MBP encoding nucleotide sequences. Useful variants of such proteins may be provided using site directed mutagenesis, eg using PCR, as is well known in the art.

As stated previously nucleic acid including that of SEQ ID No 1 encoding for DP of SEQ ID No 2 has been deposited on August 17th 1999 under the terms of the Budapest Treaty for the International Recognition of Microorganism Deposits for Patent Purposes of 28th April 1977 at the Coleccion Espanola de Cultivos Tipo in plasmid pCLON33 under deposit number CECT 5195. Restriction enzymes contained in the multicloning site of the plasmid can be used to excise the insert cDNA from this, eg. EcoRI and XhoI or BamHI and XhoI. For in vitro transcription-translation, the full-length TmDP DNA was cloned into pBluescriptSK+ using BamHI and XhoI restriction enzymes.

A nucleic acid of SEQ ID No 3 encoding for a plant E2F of amino acid sequence of SEQ ID No 3 has been deposited on 12th May 1998 under the terms of the Budapest Treaty for the International Recognition of Microorganism Deposits for Patent Purposes of 28th April 1977 at the Coleccion Espanola de Cultivos Tipo in plasmid pCLON35 under deposit number CECT5043. BamHI and XhoI, can be used to excise the insert cDNA from this. For in vitro transcription-translation, the full-length TmE2F cDNA was cloned into pBluescriptSK+ using these enzymes.

Suitable DNAs encoding for fusion proteins for use as binding materials of the invention provide a third aspect of the present invention.

For the purpose of using DNAs isolated and/or identified by the method of the present invention to detect similar genes in genomes other than that of plants, or even in plants other than wheat, it is possible to use computational means such as by

carrying out homology searches using eg. GENBANK or EMBL databases. Alternatively the DNAs of the genes so identified may be used as hybridization probes in low stringency and, preferably, high stringency probing of libraries of genomic DNA.

DNA or RNA provided from a plant or the deposit referred to above may be altered by mutagenic means such as the use of mutagenic polymerase chain reaction primers. Methods of producing the proteins or peptides of the invention characterised in that they comprise use of the DNA or RNA as described in the copending patents to express them from cells are also provided in this aspect. It may be preferred to express truncates of E2F rather than whole ORFs in order to increase expression. For example, a DNA encoding a C-terminal truncate, eg lacking from 50 to 100 amino acids, particularly lacking the 80 C-terminal amino acids from SEQ ID No 4, expresses at higher level than the whole protein in E coli.

It will be seen that it may be better to express the DP hetero-dimerization domain peptide or protein and E2F heterodimerization domain and DNA binding domain peptide or protein, even in labelled form, as a hetero-dimer from the same cell transformed with appropriate recombinant DNA. It may be still more preferred to express this with the Rb peptide or protein having the E2F binding domain. Such bound peptide or protein complex may be expected to express relatively well.

For the purpose of screening for plant DPs and E2Fs for use in the method of invention, a process which has heretofor been hampered due to human E2F and DP dissimilarity to plant E2F and DP, nucleic acid probes or primers comprising a double or single stranded DNA of sequence corresponding to 10 or more contiguous nucleotides taken from sequences SEQ ID No 1 and SEQ ID No 3 respectively are provided, with the proviso that they are not selected from those just encoding for the amino acid sequence that is relatively highly conserved with human DP or E2F. Reference should be made to the aforesaid incorporated copending patents for this. Such probes and primers may be used in Northern and Southern blotting and in PCR, including RT-PCR, and LCR.

The copending patent also describe antibodies capable of specifically binding with plant E2F. Antibodies capable of binding DP factor peptides or proteins of the first aspect of the present invention are exemplified below, thus enabling the identification and isolation of further peptides and proteins of the invention and nucleic acid sequences encoding therefor, eg. using techniques such as Western blotting.

The present invention will now be illustrated further by reference to the following non-limiting Examples and Figures. Further embodiments falling within the scope of the claims attached hereto will occur to those skilled in the art in the light of these.

FIGURES

Figure 1 shows the sequence of TmDP (wheat DP) and the amino acid sequence encoded thereby.

Figure 2 shows an alignment of TmDP protein amino acid sequence with avaialble sequences of DP proteins of animal origin. Asterisk (*) indicates amino acid identity between all sequences.

Figure 3 shows an alignment of partial sequences from the TmDP sequence and a deduced amino acid sequence corresponding to an Arabidopsis thaliana genomic DNA sequence of previously unknown function.

SEQUENCE LISTING

SEQ ID No 1 shows the nucleic acid sequence of a DNA encoding for Triticum monococcum (wheat) DP.

SEQ ID No 2 shows the amino acid sequence of Triticum monococcum (wheat) DP including leader/signal peptide sequence.

SEQ ID No 3 shows the nucleic acid sequence of a DNA encoding for Triticum monococcum (wheat) E2F.

SEQ ID No 4 shows the amino acid sequence of Triticum monococcum (wheat) E2F.

SEQ ID Nos 5, 7, 9, 11 and 13 show the nucleic acid sequences of DNAs encoding for characteristic parts of the DP protein provided by the invention.

SEQ ID Nos 6, 8, 10, 12 and 14 show the amino acid sequences of the characteristic parts of the DP protein provided by theinvention.

SEQ ID No 15 shows the nucleic acid sequence of the sense strand of double stranded DNA containing a canonical wild type E2F binding sequence.

SEQ ID No 16 shows the nucleic acid sequence of the sense strand of double stranded DNA containing a non-binding mutant E2F canonical binding sequence.

GENERAL METHODS AND PROCEDURES

Experimental procedures

Wheat cell cultures

The *Triticum monococcum* suspension culture (P.M. Mullineaux; John Innes Centre, UK), was maintained as described (Xie etal., 1995).

DNA manipulations and plasmid constructions

Standard DNA manipulation techniqueswere applied as described (Sambrook et al., 1989). DNA sequencing was carried out using an Applied Biosystem 373A device. Oligonucleotides were from Isogen Bioscience BV (Maarsen, The Netherlands).

Plasmids pGADTmE2F, pGADTmE2F(1-373),pGADTmE2F(236-458), pGADTmE2F(236-373) and pGADTmE2F(391-458) were constructed as described previously (Ramirez-Parra et al., 1999). For *in vitro* transcription-translation, the full-length TmE2F and TmDP cDNAs were cloned into pBluescriptSK+ (pBS^{SK+}-TmDP). Plasmids pGADE2F-1, pGADE2F-5, pACT2-DP1 and pACT2-DP2

containing human E2F-1, E2F-5, DP1 and DP2, respectively, were provided by N. LaThangue and S. dela Luna.

Plasmid pGBT-TmE2F was made by cloning the TmE2F cDNA in frame into the pGBT8 vector (Clontech), pGBT-TmE2F(1-373) by deleting the SspI-XhoI fragment of pGBT-TmE2F and pGBT-TmDP by cloning the TmDP cDNA in frame into the pGBT8 vector.

Plasmid pGEX-TmE2F(1-373) was constructed bycloning the SmaI-SspI fragment from pGADTmE2F plasmid in-frame into pGEX-KGvector (Pharmacia) and pMBP-TmDP by cloning the TmDP cDNA in frame into the pMal-c2 vector (New England Biolabs). Plasmid pGFP-TmDP contains SmaI-SacI fragment from pBS^{SK+}-TmDP cloned in-frame into pAVA318 vector (von Arnim et al., 1998). Construction of wheat cDNA library and plasmid DNA preparation were carried out as described in Ramirez-Parra et al., 1999.

Production and purification ofrecombinant TmE2F and TmDP proteins

E. coli BL21(DE3) transformed with plasmids expressing the GST-TmE2F and MBP-DP fusion proteins were grown toan OD_{600} of 0.6-0.9 and induced with 1 mM IPTG. GST-TmE2F was purified using glutathione-Sepharose beads(Pharmacia) while MBP-DP was purified using maltose agarose beads (NewEngland Biolabs).

For the pull-down experiments, the full-length TmE2F cDNA was *in vitro* transcribed and translated in the presence of ^{35}S-methionine using the TNT kit (Promega). *In vitro* binding experiments were carried out essentially as described (Huntley et al., 1998). The generation of the polyclonal serum against TmE2F using purified GST-TmE2F (236-458) has been described in Ramirez-Parra et al.,1999.

Electrophoretic mobility shiht assays(EMSA)

Protein extracts for DNA binding studies were prepared essentially as described in (Bogre et al., 1997). A typical binding reaction mixturecontained 20 mM Hepes, pH7.9, 12% glycerol, 50 mM KCl, 1 mM DTT, 1 mM EDTA,1 mM $MgCl_2$, 1 μg of salmon sperm and 10 μg of protein extract or 200ng of each bacterially purified MBP-TmDP and GST-TmE2F(1-373) proteins, as indicated. The binding mixture was incubated 20 minutes at 4°C and the DNA-protein complexes were fractionated by electrophoresis though 4% polyacrylamide gels at 4°C in 0.5xTBE buffer. Synthetic oligonucleotides indicated in the corresponding figure legend wereend-labeled with γ-^{32}P-ATP (top strand), annealed with an excess of the cold complementary bottom strand and used as a probe in the binding reactions. The same oligonucleotides without labelling were used as cold competitors. For the supershift assays, 2 μl of the polyclonal serum against TmE2F as described ibid were added to the binding mixture and the incubation proceeded for 10 minutes at4°C.

Yeast two-hybrid screening and assays

Yeast growth conditions and two-hybrid analysis have been described (Fields and Song, 1989; Ramirez-Parra et al., 1999). Yeasts were first transformed with plasmid pGBT-TmE2F(1-373) and, then, with the wheat cDNA library (Xie et al., 1999; Ramirez-Parra et al., 1999). The transformation mixture was plated on yeast drop-outselection media lacking tryptophan, leucine and histidine, supplemented with 10 mM 3-amino-1,2,4-triazole (3-AT). Transformants recovered during a 3-8 days period were checked for growth in the presence of ^{3}20-30 mM 3-AT. The interaction was corroborated by aβ-galactosidase assay (Breeden and Nasmyth, 1985).

RNA extraction and northern blot analysis.

Total RNA from wheat cells, leaves and roots was prepared essentially as described (Xie et al., 1999). The RNA sample (10 µg) was denatured, fractionated in a 1.2% agarose gel plus 2.2 M formaldehyde, and transferred to a Zeta-Probe membrane (Bio-Rad). The full-length TmDP probe was labelled by random priming with α-^{32}P-dCTP.

Transfection by particle bombardment and fluorescence studies.

Transient expression assays in onion epidermal cells using particle bombardment were carried out as described in von Arnim and Deng, (1994) with DNA-coated gold particles (1 µm) using a Biolistic PDS-1000/He System (BioRad). Coating of gold particles was carried out essentially as described (Sanford et al., 1993;Suarez-Lopez and Gutierrez, 1997) using 2 µg of the GFP- or the GFP-TmDP-expressing plasmids per assay. After bombardment, onion inner epidermal peels were incubated as described (Varagona et al., 1992) for 16 hr at 26°C and observed under phase contrast in a Zeiss Axiovert 35 microscope. GFP fluorescence was examinated using 480 nm excitation light and a 510 nm long-pass filter. Phase contrast and fluorescence images were photographed, digitized and processed using Adobe Photoshop software.

EXAMPLES

EXAMPLE 1.

Isolation of TmDP cDNA clones

In order to isolate protein partners for wheat TmE2F, E2F expressed from cDNA described in copending PCT/EP99/03158 was fused to the Gal4-DNA binding domain and applied to yeast two-hybrid screening. The TmE2F clone by itself trans-activated the reporter genes and could not be used as bait. Therefore the last 85 amino acids of TmE2F which, based on amino acid homology studies and binding experiments, should contain the Rb-binding and trans-activation domains but not a

putative dimerization domain, were deleted. This C-terminally truncated TmE2F(1-373) does not trans-activate the reporter genes in yeast and was used as a bait in a screening of a wheat cDNA library constructed as a fusion to the Gal4-activation domain.

The positive clones which allowed growth of yeast colonies in selective medium (-trp, -leu, -his) were tested for growth at different concentrations of 3-AT (to detect false positive in growth) and for β-galactosidase activity. Thirty clones were isolated that grew in the presence of 30 mM 3-AT and gave a strong β-galactosidase signal. Partial DNA sequencing led us to identify and isolate 6 clones, with identical DNA sequence. When we used this sequence as a query in a BLAST search against GenBank database the already identified members of the animal DP family were retrieved.

The TmDP cDNA clone isolated by two-hybrid screen is 1089 bp long, including the poly-A tract. It contains a 5'-untranslated region and a 233 bp 3'-untranslated region. The cDNA clone contains a single open reading frame of 261 amino acids (see Figure 1).

An amino acid homology study using the CLUSTALW routine (carried out on July 25, 1999) with the available sequences of DP proteins from animal origin was carried out (Figure 2). Alignment of the TmDP with the animal DP sequences available in public databases revealed the existence of several conserved motifs, strongly suggesting that the TmDP cDNA clone encodes a protein belonging to the family. This together with its ability to interact with a plant E2F protein, indicates that the TmDP cDNA encodes a bona-fide plant DP protein.

The CLUSTALW routine had parameters as follows

Program:	:	Alignment:	OutputFormat:	OutOrder:		
ClustalW		-align	aln	aligned	ktup: 1	
Window: 0	Score:	TopDiag:	PairGap: 0.05	Matrix:	GapOpen:	
	percent			blosum	10	

EndGaps: GapExt: GapDist:

10 0.05 0.05

Use JalView:

A search using the tfasta routine against the Arabidopsis thaliana database (carried out on July 23, 1999) did not retrieve a sequence with significant homology.

A search using the tblastn routine against the Arabidopsis thaliana database, including EST and BAC sequences (carried out on July 25, 1999) retrieves the information shown in Figure 3. This does not at first seem to correspond to an identified Arabidopsis protein as amino acids highly conserved in animal DPs are not contained in the sequence retrieved.

Blast parameters were set as follows

Summary of BLAST to Arabidopsis GenBank DNA sequences using a minimum match cutoff of 50%. When many high scoring pairs (HSP) are found only a subset of the best HSPs are presented. Therefore, we can conclude that, at that date, an Arabidopsis homologue of animal DP has not been identified by the genome sequencing program or by other approaches.

EXAMPLE 2.

Construction of plasmids encoding for fusion proteins.

Glutathione-S-transferase (GST) and Maltose binding protein (MBP) fusions are obtained as follows. Plasmid pGST-TmE2F(1-373) is conveniently constructed by cloning the fragment SmaI-SspI of TmE2F from pBS-TmE2F (pCLON35, CECT deposit #5043) in frame into the SmaI site of pGEX-KG vector (Pharmacia).

As an alternative plasmid pGST-TmE2F(1-373) was constructed by cloning the fragment SmaI-NcoI of TmE2F from pGBT-TmE2F(1-373) in frame into the SmaI-NcoI sites of pGEX-KG vector (Pharmacia). pGBT-TmE2F(1-373) was constructed by cloning the fragment SmaI-SspI of TmE2F from pGAD-TmE2F(1-373) (Ramirez-Parra et al., 1999; see copending patent PCT/EP99/03158) into the SmaI site of pGBT8 vector. This is a truncated version of TmE2F lacking the C-

terminal 85 aa (the overlapping trans-activation and Rb-binding domains) but retaining the dimerization and DNA-binding domain.

Plasmid pMBP-TmDP is conveniently constructed by cloning the full-lengh fragment SmaI-XhoI of TmDP from pBS-TmDP (pCLON33, CECT deposit #5195) in frame into the BamHI(blunted end)-SalI sites of of pMalc2 vector (New England Biolabs).

As an alternatively plasmid pMBP-TmDP was constructed by cloning the full-length fragment SmaI-XhoI of TmDP(1-261) from pGAD-TmDP (originally isolated clone in the two-hybrid screening) in frame into the BamHI (blunted end)-SalI sites of of pMalc2 vector (New England Biolabs). Sequence and frame were corroborated by sequencing.

EXAMPLE 3.

Protein expression in E. coli and purification

GST-TmE2F(1-373) and MBP-TmDP were expressed in E. coli BL21(DE3). Transformants were grown to an OD600 of 0.6 and induced with 1 mM IPTG for 3hours at 25°C. GST and MBP fusion proteins were purified using glutathione-Sepharose beads (Pharmacia) or Amylose Resin (New England Biolabs) respectively.

Ability to specifically bind E2F responsive DNA was assessed using electrophoretic mobility shift assays (EMSA) as follows

Gel shift reactions were carried out with 50ng of GST-TmE2F(1-373) and 100ng of MBP-TmDP in binding buffer (12% glycerol, 20mM TrisHCl pH 7.8,50mM KCl, 1mM EDTA, 1mM DTT, 1mM MgCl2) in the presence of 1μg of salmon sperm DNA. After 15 min. of incubation (to allow hetero-dimer formation), 4 fmoles (5000 cpm) of 32P-labelled ds-oligonucleotide containing the consensus sequence for human E2F-1 binding was added for 20 minutes at 4°C. Reactions were run in 4% Acrylamide gel, 0.5X TBE buffer at 4°C at 150V to separate free and complexed DNA.

Ds-oligonucleotides: These are effectively a TmE2F binding probe and a control probe. The E2F Consensus Oligonucleotide sequences being:

E2F Wild type consensus ds-oligonucleotide

5' ATT TAA GTT TCG CGC CCT TTC TCA A 3' E2F WT-Sense

3' TAA ATT CAA AGC GCG GGA AAG AGT T 5' E2F WT-Antisense

E2F Mutant ds-oligonucleotide (non-binding control)

5'ATT TAA GTT TCG ATC CCT TTC TCA A 3' E2F Mut-Sense

3'TAA ATT CAA AGC TAG GGA AAG AGT T 5' E2F Mut-Antisense

The binding sites are underlined. The sense oligonucleotide was 32P-labelled and then annealed to a sufficient excess of cold antisense oligonucleotide to ensure that labelled, free oligonucleotide is undetectable.

EXAMPLE 4.

Preparation of column capable of specifically binding transcription factor target DNA.

Glutathione SepharoseTM 4B is obtained as Catalog # 17-0756-01 (Amersham Pharmacia Biotech AB). Amylose resin is obtained as Catalog # 800-215 (New England Biolabs). Buffer conditions are as used for EMSA binding buffer with the modification that glycerol concentration may be reduced to 6% if desired. ie. 6% glycerol, 20mM TrisHCl pH 7.8, 50mM KCl, 1mM EDTA, 1mM DTT, 1mM MgCl2.

The GST-E2F-containing bacterial protein extract is applied to the Glutathione Sepharose resin, according to the manufacturer's instructions using the buffer described above. While MBP-DP-containing bacterial protein extract is applied to the Amylose resin, according to the manufacturer's instructions, using that same buffer. Proteins are eluted as pure according to maufacturer's instructions and equimolar amounts of both are mixed at 4°C for 1 hour.

The mixture is applied to through a new Glutathione Sepharose resin, equilibrated in the same buffer to purify the hetero-dimer, after washing according to

the manufacturer's instructions. NB: the mixture can be passed through an amylose resin to retain the hetero-dimer via the MBP moiety. The hetero-dimer bound to the resin is now ready to be used to retain the pretreated genomic DNA. Pretreatment of this DNA is by shearing or by digestion with Sau3A to provide fragments.

EXAMPLE 5

Purified plant E2F binds DNA with arelatively low efficiency

In order to study in more detail E2F-DNA protein complex formation we carried out EMSA with the bacterially-expressed and purified GST-TmE2F fusion protein which contained the deletion of the last 85 amino acids, but still retained the conserved DNA-binding domain (this truncated version was expressed in much larger amounts than the full-length protein).

A single DNA-protein complex was detected with the wild type E2 binding site probe, when increasing amounts of purified TmE2F were added to the binding mixture. The formation of this TmE2F-DNA complex was specific since (i) complex formation depends on an intact E2F binding site as it did not occur when the E2 mutant probe, containing the point mutations within the canonical E2F binding site, was used, and (ii) when purifed GST proteins was added. To confirm these observations we also carried out competition experiments. Adding increasing molar excessof the E2 wild type probe, but not of the mutated E2 mutant probe, was able to compete out the preformed TmE2F-DNA complexes. Altogether, these data indicate that TmE2F binds effectively to one of the DNA sequences described as a canonical E2F DNA-binding sequence for human E2F-1 and that this binding was specific and depends on an intact E2F binding site.

It should be pointed out, however, that in these binding studies the amount of purified E2F protein needed to efficiently form a complex was relatively high, suggesting that complex formation was not optimal, even under a wide variety of binding conditions (not shown).

EXAMPLE 6.

Heterodimerization properties of TmE2F and TmDP

It has been previously shown that human DP-1 and DP-2 can form stable heterodimer with any of the E2F family members (Bandara et al., 1993; Helinet al., 1993; Sardet et al., 1995; Ormondroyd et al., 1995), a situation which is different from the more stringent interaction observed between E2F and pocket proteins (Dyson,1998; Ramirez-Parra et al., 1999). To investigate the heterodimerization properties of TmDP the inventors used a yeast two-hybrid approach using E2F and DPproteins of different sources. All the combinations tested between human (HuE2F-1,HuE2F-5, HuDP-1 and HuDP-2) and plant (TmE2F and TmDP) proteins allowed efficient growth of the cotransformant in selective medium, supplemented with 20 mM 3-AT and were positive in the β-galactosidase assay. Therefore, we can conclude that the protein domains involved in heterodimerization are functionally conserved between organisms as divergent as human and wheat, a result consistent with the high amino acid conservation observed in their heterodimerization domains. Deletion experiments confirmedthat a central domain in TmE2F, which contains the leucine zipper, is necessary and sufficient to mediate the heterodimerization with TmDP.

EXAMPLE 7.

TmDP stimulates binding ofTmE2F to a canonical DNA binding site

To determine whether TmDP had any functional effect on the TmE2F DNA binding activity, theinventors carried out EMSA with purified proteins. Addition of purified MBP-TmDP to a DNA probe containing a canonical E2F binding site (E2wild type oligo) did not produce any retarded band, indicating that TmDP alone does not bind to DNA. However, under conditions of low amounts of TmE2F, where DNA binding was virtually undetectable, adding increasing amounts of purified MBP-TmDP very significantly stimulated complex formation. Therefore, it can be

concluded that heterodimerization of TmE2F with TmDP contributes to a several-fold increase in the affinity and/or stability of the DNA-protein complex.

EXAMPLE 8.

Demonstration that TmDP is a cytoplasmic protein

To investigate the subcellular localization of TmDP we used transient expression assays after biolistic delivery to onion epidermal cells (Varagona et al., 1992) of a translational fusion of TmDP to the greenfluorescent protein (GFP; Sheen et al., 1995) under the control of the CaMV 35S promoter, which allowed us to transiently follow the expression of the chimeric construct. In all cases observed, the TmDP-GFP fusion proteinwas present in the cytoplasm of transfected onion cells and apparently excluded from the nucleus, a pattern which was different from that of the control cells expressing GFP alone which is known to also diffuse to the nucleus. This observation is consistent with the lack of a nuclear localization signal (NLS) in the TmDP amino acid sequence. Therefore, it is most likely that the putative NLS identified in TmE2F is responsible for actively transporting the TmDP-TmE2F heterodimer to the nucleus.

To demonstrate that TmDP alone does not have transactivation potential in yeast the inventors transformed yeast cells with a plasmid expressing TmDP fused to the Gal4 DNa-binding domain andplated them under selective conditions (-trp, ±his) in the presence of 20 mM 3-AT. Yeast cells were unable to grow underthose conditions, as they were those carrying the vector alone, while yeast cells expressing TmE2F fused to the Gal4 DNA binding domain could grow, indicating that TmDP lacks transactivation potential on its own. In the same assay, human DP-1 and DP-2 alone do not transactivate either. Therefore, it is likely that thetransactivation ability of the TmE2F/TmDP heterodimer is conferred by the C-terminal domain of TmE2F

in theE2F/DP heterodimer, although the presence of the DP partner may cooperatein E2F transactivation by stabilizing the DNA-protein complex.

EXAMPLE 9.

Characterisation of TmDP

The idea that the isolated CDNA encodes a plant member of the DP family was reinforced by analysis of the amino acid homology and domain organization. TmDP exhibits an overall 29-33% aminoacid similarity with human (Bandara et al., 1993; Girling et al., 1993;Krek et al., 1993) and X. laevis (Girling et al., 1994) DP-1 and DP-2 and a slightly smaller similarity(27%) with D. melanogaster DP (Dynlacht et al., 1994; Ohtani and Nevins, 1994). Amino acid alignment of plant and animal DP proteins indicates that it has a similar domain organization (Fig. 3B). The highest homology occurs within a 70 amino acidregion (residues 64-143 in TmDP) which in animal DP proteins are important for DNA binding (Wu et al., 1996). This region includes a 10 amino acid stretch of fully conserved residues. Other amino acid blocks with a significant degree of homology contain the heptad repeats (residues 144-213 in TmDP), involved in heterodimerization with E2F (Wu et al., 1996; Zheng et al., 1999) and the domain conserved with E2F proteins (residues 214-240 in TmDP), a region which is similar to the E2F family members (Girling et al., 1993). Experimental evidences of the heterodimerization properties of TmDPwill be presented below in detail. Quite interestingly, TmDP lacks an acidic region which is present near the C-terminus of animal DP members, a domain whose functional significance has not been determined. Finally, the less conserved region corresponds to the N-terminal domain whose length and amino acid sequence is similar to that of animal DP members, in particular to the DP-2 group. Based on these homology studies, we conclude that TmDP presents a higher amino acid sequence similarity to animal DP-2. However, it is worth noting its smaller size and the absence of an acidic C-terminal domain as unique properties of TmDP.

To confirm that the cDNA isolated contained a single ORF *in vitro* transcription and translation reactions of a suitable construct were carried out in the

presence of ^{35}S-labelled methionine. A major ~30 kDa band was obtained, consistent with the predicted molecular mass. Finally, to determine whether TmE2F/TmDP interaction can occur in the absence of other proteins, we carried out in vitro pull-down experiments. When *in vitro* translated ^{35}S-labelled TmE2F was mixed with bacterially expressed and purified MBP-TmDP, but not with MBP alone, a significant amount of the input material remained bound to the resin indicative of a direct interaction between TmE2F and TmDP.

Northern analysis of total RNA samples reveals the existence of a major ~1.9±0.2 kb transcript together with another much larger transcript (3.5±0.2 kb). Both transcripts were present in cultured cells, where they were more abundant, as well as in roots and leaves. However, the smaller transcript was more abundant in leaves. The larger transcript may represent a partially processed RNA species or the result of an alternative splicing. It is important to mention that alternative splicing is characteristic of mammalian DPproteins (Ormondroyd et al., 1995; de la Luna et al., 1996; Rogers et al., 1996; Wu et al., 1996), although its functional significance has not been established yet.

The murine homolog of human DP-2 (DP-3) is unique in that its primary transcript undergoes extensive alternative splicing giving rise to a final complex mixture of four products (α, β, γand δ; Ormondroyd et al., 1995). The α and δ proteins share a common alternative spliced exon, encoding 16 amino acid residues, known as the E region, which is absent in the β and γ isoforms as well as in DP-1 and which seems to function as a nuclear localization signal (NLS; de la Luna et al., 1996). The β, γ and δ isoforms are produced after initiation of translation in a methionine residue downstream from the first methionine, used in the α isoform. In addition, the γ isoform has an extra glutamine residue. Thus, based on the amino acid sequence of TmDP, the presence of a relatively short N-terminal region, the lack of an apparent E region, and consequently of a NLS, and the lack of the Q residue within the DNA-binding domain, TmDP seems to be structurally related to the murine DP-2β isoform (de la Luna et al., 1996). It is interesting to note that the small size of

TmDP does not seem to be caused by a lack of N-terminal residues but, rather, to a lack of a large C-terminal region, which is present in animal DP proteins. In the latter, this region is highly acidic but its function has not been established yet (Wu et al., 1995; Zhang and Chellapan, 1995).

While this work was in process, the sequence of genomic regions of A.thaliana encoding putative DP-like proteins has been released. In the absence, sofar, of sequence information of the corresponding cDNAs, it is difficult to make a detailed homology study due to potential uncertainties inherent to the amino acid predictions of the intron/exon sequences. An additional complication in the case of DP-like proteins may derive from the possibility of differentially spliced forms of plant DP transcripts. As in the case of TmDP, the predicted sequences of A. thaliana DP-like proteins have a relatively short N-terminus and lack the extended C-terminal end characteristic of the animal DP proteins.

REFERENCES

Ach, R.A., Durfee, T., Miller,A.B., Zambryski, P.C., Hanley-Bowdoin, L. and Gruissem, W. (1997a) *RRB1* and *RRB2* encode maize retinoblastoma-related proteins that interact with a plantD-type cyclin and geminivirus replication protein. *Mol. Cell. Biol.***17**, 5077-5086.

Albani, D., Mariconti, L., Ricagno, S., Pitto, L., Moroni, C., Helin,K. and Cella, R. (2000) DcE2F, a Functional Plant E2F-like TranscriptionalActivator from Daucus carota. *J. Biol. Chem.* **275**, 19258-19267.

Bandara, L.R., Buck, V.M., Zamanian, M., Johnston, L.H. and LaThangue,N.B. (1993) Functional synergy between DP-1 and E2F-1 in the cellcycle-regulating transcription factor DRTF1/E2F. *EMBO J.* **12**, 4317-4324.

Bogre et al., 1997, Plant Cell, 9: 75-83

Breeden, L. and Nasmyth, K. (1985) Regulation of the yeast HO gene.*Cold Spring Harbor Symp. Quant. Biol.* **50**, 643-650.

Breeden, L. (1996) Start-specific Transcription in Yeast. *Curr.Topics Microbiol. Immunol.* **208**, 95-127.

de Jager, S.M. and Murray, J.A.H. (1999) Retinoblastoma proteins inplants. *Plant Mol. Biol.* **41**, 295-299.

de la Luna, S., Burden, M.J., Lee, C.W. and LaThangue, N.B. (1996)Nuclear accumulation of the E2F heterodimer regulated by subunitcomposition and alternative splicing of a nuclear localization signal. *J. Cell Sci.* **109**, 2443-2452.

Doonan, J. and Fobert, P. (1997) Conserved and novel regulators of theplant cell cycle. *Curr. Opin. Cell Biol.* **9**, 824-830.

Dynlacht, B.D., Brook, A., Dembski, M., Yenush, L. and Dyson, N.(1994) DNA-binding and trans-activation properties of Drosophila E2F and DPproteins. *Proc. Natl. Acad. Sci. USA* **91**, 6359-6363.

Dyson, N. (1998) The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245-2262.

Fields, S. and Song, O. (1989) A novel genetic system to detectprotein-protein interactions. *Nature* **340**, 245-246.

Fountain, M.D., Murray, J.A.H. and Beck, E. (1999) Isolation andcharacterization of a plant retinoblastoma-related gene fromphotoautotrophic cell suspension culture of *Chenopodium rubrum L. Plant Physiol.* **119**, 363.

Fuerst, R.A., Soni, R, Murray, J.A.H. and Lindsey, K. (1996)Modulation of cyclin transcript levels in cultured cells of *Arabidopsisthaliana. Plant Physiol.* **112**, 1023–1033.

Girling, R., Partridge, J.F., Bandara, L.R., Burden, N., Totty, N.F.,Hsuan, J.J. and La Thangue, N.B. (1993) A new component of transcriptionfactor DRTF1/E2F. *Nature* **362**, 83-87.

Girling et al., 1994

Grafi, G., Burnett, R.J., Helentjaris,T., Larkins, B.A., DeCaprio,J.A., Sellers, W.R. and Kaelin, W.G.Jr. (1996) A maize cDNA encoding amember of the retinoblastoma protein family: involvement inendoreduplication. *Proc. Natl. Acad. Sci. USA* **93**: 8962-8967.

GutiŽrrez, C. (1998) The retinoblastoma pathway in plant cell cycleand development. *Curr. Op. Plant Biol.* **1**, 492-497.

GutiŽrrez, C. (2000) DNA replication and cell cycle in plants:learning from geminiviruses. *EMBO J.* **19**, 792-799.

Hao et al., 1995

Helin, K., Lees, J.A., Vidal, M., Dyson, N., Harlow, E. and Fattaey,A. (1992) A cDNA encoding a pRB-binding protein with properties of thetranscription factor E2F. *Cell* **70**, 337-350.

Helin, K., Wu, C.L., Fattaey, A.R., Lees, J.A., Dynlacht, B.D., Ngwu,C. and Harlow, E. (1993a) Heterodimerization of the transcription factorsE2F-1 and DP-1 leads to cooperative trans-activation. *Genes Dev.* **7**, 1850-1861.

Helin, K., Harlow, E. and Fattaey, A. (1993b) Inhibition of E2F-1trans-activation by direct binding of the retinoblastoma protein. *Mol.Cell. Biol.* **13**, 6501-6508.

Helin et al., 1997

Helin, 1998

Helin, 2000

Huntley, R.P.A. and Murray, J.A.H. (1999) The plant cell cycle. *Curr. Opin. Plant Biol.* **2**, 440-446.

Huntley, R., Healy, S., Freeman, D., Lavender, P., de Jager, S.,Greenwood, J., Makkerh, J., Walker, E., Jackman, M., Xie, Q., Bannister,A.J., Kouzarides,

T., Gutierrez, C., Doonan, J.H. and Murray, J.A.H. (1998) The maize retinoblastoma protein homologue ZmRb1 is regulatedduring leaf development and displays conserved interactions with G1/Sregulators and plant cyclins (D-type) proteins. *Plant Mol. Biol.* **37**, 155-169.

Kosugi, S. and Ohashi, Y. (1997) PCF1 and PCF2 Specifically bind tocis Elements in the Rice Proliferating Cell Nuclear Antigen Gene. *PlantCell.* **9**, 1607-1619.

Kozak, M. (1987) At least six nucleotides preceding the AUG initiatorcodon enhace translation in mammalian cell. *J. Mol. Acad. Sci. USA.***87**, 6922-6926.

Krek, W., Livingston, D.M. and Shirodkar, S. (1993) Binding to DNA andthe retinoblastoma gene product promoted by complex formation of differentE2F family members. *Science* **262**, 1557-1560.

Lowndes, N.F., McInerny, C.J., Johnson, A.L. Fantes, P.A. and Johnston,L.H. (1992) Control of DNA synthesis genes in fission yeast by the cellcycle gene cdc10. *Nature* **355**, 449-453.

Lui, H., Wang, H., DeLong, C., Fowke, L.C., Crosby, W.L. and Fobert,P.R. (2000) The Arabidopsis Cdc2a-interacting protein ICK2 is structurallyrelated to ICK1 and is a potent inhibitor of cyclin-dependent kinaseactivity in vitro. *Plant J.* **21**, 379-386.

Mart'nez-Balb‡s, M.A., Bauer, U.M., Nielsen, S.J., Brehm, A. andKouzarides, T. (2000) Regulation of E2F1 activity by acetylation. *EMBOJ.* **19**, 662-671.

Marzio, G., Wagener, C., Gutierrez, M.I., Cartwright, P., Helin, K. andGiacca, M. (2000) E2F family members are differentially regulated byreversible acetylation. *J.Biol. Chem.* **275**, 10887-10892.

Mironov, V., De Veylder, L., Van Montagu, M. and InzŽ, D. (1999)Cyclin-dependent kinases and cell division in plants - the Nexus. *PlantCell* **11**, 509–521.

Nakagami, H., Sekine, M., Murakami, H. and Shinmyo, A. (1999) Tobaccoretinoblastoma-related protein phosphorylated by a distinctcyclin-dependent kinase complex with Cdc2/cyclin D in vitro. *Plant J.***18**, 243–252.

Nevins, J.R. (1992) E2F: a link between the Rb tumor suppressorprotein and viral oncoproteins. *Science* **258**, 424-429.

Ohtani, K. and Nevins, J.R. (1994) Functional properties of aDrosophila homolog of the E2F-1 gene. *Mol. Cell. Biol.* **14**,1603-1612.

Ormondroyd, E., de la Luna, S. and La Thangue, N.B. (1995) A newmember of the DP family, DP-3, with distinct protein products suggests aregulatory role for alternative splicing in the cell cycle transcriptionfactor DRTF1/E2F. *Oncogene* **11**, 1437-1446.

Philips et al., 1995

Ram'rez-Parra, E., Xie, Q., Boniotti M.B. and Gutierrez, C. (1999)The cloning of plant E2F, a retinoblastoma-binding protein, reveals uniqueand conserved features with animal G1/S regulators. *Nucleic Acids Res.***27**, 3527–3533.

ref 4 in albani

Riou-Khamlichi, C., Huntley, R., Jacqmard, A. and Murray, J.A.H. (1999)Cytokinin activation of Arabidopsis cell division through a D-type cyclin.*Science* **283**, 1541–1544.

Rogers et al., 1996

Sambrook, J., Fritsch, E.F. and Maniatis,T. (1989) Molecular cloning:A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold SpringHarbor, NY, USA.

Sanford, J.C., Smith, F.D. and Russell, J.A. (1993) Optimizing thebiolistic process for different biological applications. *MethodsEnzymol.* **217**, 483-509.

Sardet, C., Vidal, M., Cobrinik, D., Geng, Y., Onufryk, C., Chen, A.and Weinberg, R.A. (1995) E2F-4 and E2F-5, two members of the E2F family,are expressed in the early phases of the cell cycle. *Proc. Natl. Acad.Sci. USA* **92**, 2403-2407.

Sauter, 1997

Scott, A, Wyatt, S. , Tsov, P.L., Roberso, D. and Allen, N.S. (1999)Model system for plant cell biology: GFP imaging in living onion epidermalcells. *Biotechniques* **26**, 1125-1132.

Sekine, M., Ito, M., Uemukai, K., Maeda, Y., Nakagami,H. and Shinmyo,A. (1999) Isolation and characterization of the E2F-like gene in plants.*FEBS Lett.* **460**, 117-122.

Sheen et al., 1995

Shen and Gigot, 1997

Slansky, J.E. and Farnham, P.J. (1996) Introduction to the E2F family:protein structure and gene regulation. *Curr. Topics Microbiol. Immunol.***208**, 1-30.

Suarez-Lopez, P. and Gutierrez, C. (1997) DNA replication of wheatdwarf geminivirus vectors: effects of origin structure and size. *Virology* **227**, 389-399.

Umeda, M., Bhalearao, R.P., Schell, J., Uchimiya, H. and Koncz, C.(1998) A distinct cyclin-dependent kinase–activating kinase ofArabidopsis thaliana. *Proc. Natl. Acad. Sci. USA* **95**, 5021-5026.

Varagona, M. J., Schmidt, R. J. and Raikhel, N. V. (1992) Nuclearlocalization signal(s) required for nuclear targeting of the maizeregulatory protein Opaque-2. *Plant Cell* **4**, 1213-1227.

von Arnim, A.G. and Deng, X.W. (1994) Light inactivation ofArabidopsis Photomorphogenic Repressor COP1 Involves a Cell-SpecificRegulation of Its Nucleocytoplasmic Partitioning. *Cell.* **79**,1035-1045.

von Arnim, A.G., Deng, X.W. and Stracey, M.G. (1998) Cloning vectorsfor expression of green fluorescent protein fusion proteins in transgenicplants. *Gene* **221**, 35-43.

Wang, H., Qi, Q.G., Schorr, P., Cutler, A.J., Crosby, W.L. and Fowke,L.C. (1998) ICK1, a cyclin-dependent protein kinase inhibitor from *Arabidopsis thaliana* interacts with both cdc2a and CycD3, and itsexpression is induced by abscisic acid. *Plant J.* **15**, 501–510.

Wu et al., 1995

Wu et al., 1996

Xie, Q., Suarez-Lopez, P. and Gutierrez, C. (1995) Identification andanalysis of a retinoblastoma binding motif in the replication protein of aplant DNA virus: requirement for efficient viral DNA replication. *EMBOJ.* **14**, 4073-4082.

Xie, Q., Sanz-Burgos, A.P., Hannon, G.J. and Gutierrez, C. (1996)Plant cells contain a novel member of the retinoblastoma family of growthregulatory proteins. *EMBO J.* **15**, 4900-4908.

Xie, Q., Sanz-Burgos, A.P., Guo, H., Garcia, J.A. and Gutierrez, C.(1999) GRAB proteins, novel members of the NAC domain family, isolated bytheir interaction with a geminivirus protein. *Plant Mol. Biol.* **39**, 647-656.

Zhang and Chellapan, 1995

Zheng, N., Fraenkel, E., Pabo, C.O. and Pavletich, N.P. (1999)Structural basis of DNA recognition by the heterodimeric cell cycletranscription factor E2F-DP. *Genes Dev.* **13**, 666-674.

CLAIMS

1. A method of controlling one or more of plant growth, gene expression, cellular DNA replication, cell cycle progression, differentiation and development comprising increasing or decreasing E2F-dimerization partner (DP) protein activity in a plant cell through expression of a recombinant DP peptide or protein in that cell characterised in that the peptide or protein comprises a sequence SEQ ID No 2, a functional part thereof, or a sequence having at least 70% homology to either, that peptide or protein being capable of interacting with a plant E2F protein or peptide such as to alter E2F activity in the plant cell.

2. A method as claimed in Claim 1 characterised in that the peptide or protein sequence is of 50% or more identity with that of the corresponding full length or part of SEQ ID No 2.

3. A method as claimed in Claim 1 or Claim 2 characterised in that the peptide or protein sequence is of 70% or more identity with that of the corresponding full length or part of SEQ ID No 2.

4. A method as claimed in any one of the preceding claims characterised in that the plant DP activity comprises one or both of (i) the ability to dimerize with plant E2F protein and (ii) the ability to modulate E2F binding to E2F/DP transcription factor binding sites in plant DNA.

5. A method as claimed in any one of the preceding claims characterised in that it includes steps of altering the plant DP protein level, the E2F-DP DNA-binding activity, transactivation properties, and/or the DP/E2F-binding activity.

6. A method as claimed in any one of the preceding claims characterised in that the DP may be modified alone and/or in combination with a modification of the levels or activity of plant E2F and/or plant Rb.

7. An isolated, enriched, cell free and/or recombinantly produced protein or peptide, capable of altering E2F-dimerization partner (DP) activity in a plant cell, characterised in that it has one or both DP activities in plants selected from (i) the ability to dimerize with plant E2F protein and (ii) the ability to modulate, particularly enhance, E2F binding to E2F transcription factor binding sites in plant DNA or effect thereof

characterised in that the protein or peptide comprises an amino acid as shown in SEQ ID No 2 or a functionally active part thereof or a sequence having at least 70% homology to such sequence or part.

8. A protein or peptide as claimed in Claim 7 characterised in that it has at least 50% identity to the amino acid sequence as shown in SEQ ID No 2 or a functionall active part thereof.

9. A protein or peptide as claimed in Claim 7 or Claim 8 characterised in that it is of SEQ ID No 2 or is a variants thereof modified such that the amino acid sequence is mutated such that its ability to dimerize with E2F protein is reduced from that of the native sequence or abolished completely therefrom, whereby the peptide is capable of acting as a DP protein which decreases or abolishes native or recombinant E2F binding to its DNA binding site, thus inhibiting or abolishing E2F activity in a cell in which is its present.

10. A protein or peptide as claimed in any one of the preceding Claims 7 to 9 characterised in it comprises a sequence found in that of SEQ ID No 2 or having at least 70% homology thereto selected from those comprising

SEQ ID No 6 ARAAMAPPRGGAAAAATAALDLTGVHILEAS SVPPLPE RGGNAVQRKGAVDP

SEQ ID No 8 DKDRKKEKAAAPRITGWGLREYSKIVCEKVEAKGRT TY NEVADEIYSELKS

SEQ ID No 10 MAHIGQGFDEKNIRRRVYDAFNVLIALRVIAKEKKEIR W MGLSNYRYEKIKKLEEV

SEQ ID No 12 RKELVNKIRNKKALLQEIEKQFDDLQNIKLRNQTLESS A ENVNGIRLPFVLVKTSR

SEQ ID No 14 KARVEIEISDDSKFAHFEFNGAPFTLHDDLSILEGVRGNS IGKAGRATLH

11. A protein or peptide as Claimed in Claim 10 characterised in that the sequence comprises two or more of these sequences or sequences at least 70% homologous thereto.

12. Isolated, enriched, cell free and/or recombinant nucleic acid comprising a sequence encoding for expression of a protein or peptide as described in any one of Claims 7 to 11.

13. Nucleic acid as claimed in Claim 12 characterised in that it includes the coding nucleic acid sequence of SEQ ID No 1 or a part thereof encoding for all or a functional part of the amino acid sequence shown therein as defined above.

14. Nucleic acid as claimed in Claim 12 or Claim 13 as contained in plasmid pCLON33, deposit number CECT 5195 made on August 17th 1999 under the terms of the Budapest Treaty for the International Recognition of Microorganism Deposits for Patent Purposes of 28th April 1977 at the Coleccion Espanola de Cultivos Tipo.

15. A nucleic acids as claimed in any one of Claims 12 to 14 characterised in that it encodes for a plant DP or a functional variant thereof including the sequence of SEQ ID No 1 or a sequence complementary thereto or otherwise antisense thereto.

16. A nucleic acid probe characterised in that it comprises a DNA sequence corresponding to an amino acid sequence selected from SEQ ID No 2 to 8.

17. A nucleic acid probe or primer characterised in that it comprises a double or single stranded DNA of sequence corresponding to 10 or more contiguous nucleotides taken from the sequence SEQ ID No 1 with the proviso that it is not selected from amino acids 70 to 136. Such probes and primers may be used in Northern and Southern blotting and in PCR, including RT-PCR, and LCR.

18. An oligonucleotides probe characterised in that it comprises at least 18 contiguous bases of the SEQ ID No 1.

19. An oligonucleotide probe as claimed in Claim 18 characterised in that it is 30 to 100 bases long.

20. An oligonucleotide primer as claimed in Claim 17 characterised in that it is of 10 to 20 bases long.

21. Antisense DNA to a nucleic acids as claimed in any one of Claims 12 to 20.

22. A nucleic acid characterised in that it encodes a the DP protein or peptide encoding sequence as claimed in any one of Claims 7 to 11 together with a sequence encoding an E2F protein or peptide.

23. A nucleic acid as claimed in Claim 22 charactersied in that it the sequences encoding the DP and E2F encoding sequences are under control of the same regulatory element or elements.

24. A nucleic acid vector or construct comprising a nucleic acid as claimed in any one of Claims 12 to 23 or antisense nucleic acid thereto.

25. A plant cell comprising recombinant nucleic acid, a vector or a construct as claimed in any one of Claims 12 to 24.

26. A transgenic plant or part thereof comprising recombinant nucleic acid, a vector, DNA construct or cell as claimed in any one of Claims 12 to 25.

27. An antibody chharactersised in that it has been raised against a DP peptide or protein as claimed in any one of Claims 7 to 11.

28. An antibody as claimed in Claim 27 characterised in that it is capable of specifically binding with plant DP factor peptides or proteins as claimed in any one of Claism 7 to 11, but not to the human, mouse or Xenopus DP.

29 A method for identifying and/or isolating DNAs corresponding to complete or partial genes that are regulated in G1 passsage, G1/S-phase transition and/or S phase progression of the cell cycle, said method comprising contacting a sample of genomic DNA with a binding material specific for binding such complete or partial genes, removing non-bound DNA from the specific binding material then, releasing and isolating the bound DNA.

characterised in that the specific binding material comprises a peptide or protein including the DNA binding sequence of a protein that is capable of acting as a part of a plant hetero-oligomer transcription activator or repressor.

30. A method as claimed in Claim 29 characterised in that the specific binding material comprises a peptide or protein which includes a plant E2F DNA binding domain.

31. A method as claimed in Claim 30 characterised in that the specific binding material comprises a peptide or protein which includes a plant E2F DNA binding domain together with a plant E2F-dimerization partner (DP) hetero-dimerization domain and/or a plant retinoblastoma protein E2F binding domain.

32. A method as claimed in any one of Claims 29 to 31 characterised in that the material comprises a peptide or protein that comprises a plant E2F DNA binding domain together with a plant E2F-dimerization partner (DP) binding domain together with a peptide or protein that includes a plant dimerization partner (DP) E2F binding domain.

33. A method as claimed in any one of Claims 29 to 32 characterised in that the E2F and DP domains are comprised in Triticum monnococcum sequences or functional variants or parts thereof.

34. A method as claimed in Claim 33 charactersied in that the DP domains are comprised in peptides or proteins as claimed in any one of Claims 7 to 11.

35. A method as claimed in any one of Claims 29 to 34 characterised in that the specific binding material peptide or protein is labelled or tagged to assist in identifying or immobilising it, particularly when in bound complex with the gene to be identified and/or isolated.

36. A method as claimed in any one of Claims 29 to 35 characterised in that the specific binding material comprises two peptides or proteins, one including the E2F

DNA binding domain, and one including the DP hetero-dimerization domain bound together as a hetero-dimer.

37. A method as claimed in any one of Claims 29 to 36 characterised in that the E2F sequence is all or part of SEQ ID No 4 or a sequence having 90% homology thereto.

38. A method as claimed in any one of Claims 29 to 37 charaterised in that the binding material comprises a peptide or protein including an optionally labelled plant retinoblastoma protein E2F binding domain.

39. A specific binding material characterised in that it comprises a peptide or protein having DNA binding activity with respect to plant DNA transcription activator or repressor factor binding sites, particularly in genomic DNA, and having the ability to dimerize or oligomerize with a further such plant protein together with one or more of said further a peptides or proteins.

40. A specific binding material characterised in that it comprises a peptide or protein having DNA binding activity with respect to plant DNA E2F transcription factor binding sites, particularly those present in genomic DNA and having the ability to dimerize with plant DP protein together with one or both of

(ii) a peptide or protein that is capable of binding to plant E2F through its DP hetero-dimerization domain and

(iii) a peptide or protein that is capable of binding to plant E2F through its retinoblastoma protein binding domain.

41. A material as claimed in Claim 40 characterised in that the peptides or proteins are wheat peptides or proteins and functional variants and parts thereof.

42. A material as claimed in Claim 40 or 41 characterised in that it comprises a hetero-dimer of (i) and (ii) or (i) and (iii).

43. A material as claimed in any oen fo Claims 40 to 42 characterised in that it comprises both of the peptides or proteins (i) and (ii) in the form of a hetero-dimer together with peptide or protein (iii).

44. A material as claimed in any one of Claims 40 to 43 characterised in that the peptides or proteins are in labelled or tagged form.

45. A material as claimed in any one of Claims 40 to 44 characterised in that it is in the form of a coating or otherwise bound form on a support material.

46. A material as claimed in any one of Claims 40 to 44 charactersied in that the material is in the form of particles of the peptides or proteins and/or hetero-dimers or trimers.

47. A nucleic acid charactersied in that it encodes for a peptide or protein as claimed in any one of Claims 7 to 11 fused to a sequence encoding for a proetin label.

```
1GAATTCGGCACGAGCCGCAATGGCGCCTCCCCGCGGCGGAGCTGCTGCGGCCGCTACCGCC   61
                M  A  P  P  R  G  G  A  A  A  A  T  A          14
61  GCACTGGACCTGACCGGCGTGCACATTCTCGAAGCTTCCAGTGTCCCCCCGCTTCCCGAA 120
15    A  L  D  L  T  G  V  H  I  L  E  A  S  S  V  P  P  L  P  E  34
121 CGCGGCGGTAATGCGGTCCAAAGGAAGGGGGCTGTTGACCCGGATAAAGATAGGAAGAAG 180
35    R  G  G  N  A  V  Q  R  K  G  A  V  D  P  D  K  D  R  K  K  54
181 GAGAAGGCTGCGGCACCGAGGATCACCGGTTGGGGGCTCCGCGAGTACAGCAAAATAGTT 240
55    E  K  A  A  A  P  R  I  T  G  W  G  L  R  E  Y  S  K  I  V  74
241 TGTGAGAAAGTTGAAGCCAAAGGAAGAACAACATACAATGAGGTTGCAGACGAAATTTAT 300
75    C  E  K  V  E  A  K  G  R  T  T  Y  N  E  V  A  D  E  I  Y  94
301 TCAGAGCTGAAGTCCATGGCACATATTGGTCAAGGGTTTGATGAGAAGAATATTAGGCGG 360
95    S  E  L  K  S  M  A  H  I  G  Q  G  F  D  E  K  N  I  R  R  114
361 AGAGTGTATGATGCTTTCAACGTTCTCATTGCACTTCGTGTTATTGCAAAGAAAAAAAG 420
115   R  V  Y  D  A  F  N  V  L  I  A  L  R  V  I  A  K  E  K  K  134
421 GAGATACGGTGGATGGGCCTTTCAAATTACAGATATGAAAAATAAAGAAGCTTGAGGAA 480
135   E  I  R  W  M  G  L  S  N  Y  R  Y  E  K  I  K  K  L  E  E  154
481 GTTCGTAAAGAACTCGTCAACAAGATTAGGAACAAGAAGGCACTCCTCCAGGAAATCGAA 540
155   V  R  K  E  L  V  N  K  I  R  N  K  K  A  L  L  Q  E  I  E  174
541 AAACAGTTTGATGATCTCCAAAACATCAAGTTACGTAACCAACACTGGAAAGCTCAGCA 600
175   K  Q  F  D  D  L  Q  N  I  K  L  R  N  Q  T  L  E  S  S  A  194
601 GAGAATGTTAATGGCATCCGCCTTCCATTCGTATTGGTCAAGACATCTAGGAAAGCAAGG 660
195   E  N  V  N  G  I  R  L  P  F  V  L  V  K  T  S  R  K  A  R  214
661 GTGGAAATTGAGATTTCAGATGACTCGAAGTTTGCCCATTTCGAGTTCAATGGTGCACCA 720
215   V  E  I  E  I  S  D  D  S  K  F  A  H  F  E  F  N  G  A  P  234
721 TTCACATTGCATGATGATCTCTCAATCCTTGAGGGGGTAAGGCGTAACAGCATAGGAAGA 780
235   F  T  L  H  D  D  L  S  I  L  E  G  V  R  R  N  S  I  G  R  254
781 GCTGGCCGCGCCACCCTTCACTAGAGACTCAAGAATATTACAAATGAATTAAAAGTGTTA 840
255   A  G  R  A  T  L  H  *261
841 GAACTGGCACAGCCGGATTCTTTTGCACAGCTATGTATAGCTATATATCCTCATGAAAAC 900
901 TTGACCTAGTTTATAGGACAGTCTCTCAGGCTTGAGAAGATTTTAACCTGCAAATTTTGT 960
961 CTCCTTTTTGTGCCTAGCAGGTTATTAGGTCTCAGATAGATGATTCATATATGTGCTGCT 1020
1021 ATGAAAACATTGATAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1080
1081 AAAAAAA1089
```

FIG. 1 Sequence of TmDP cDNA and deduced amino acid sequence.

```
hDP-1   MAKDAGLIEANGELKVFIDQNLSPGKGVVSLVAVHPSTVNPLGKQLLPKTFGQSNVNIAQ 60
mDP-1   MAKDASLIEANGELKVFIDQNLSPGKGVVSLVAVHPSTVNTLGKQLLPKTFGQSNVNITQ 60
XlDP-1  MAKDAGLIEANGELKVFVDQNLSPGKGVVSLLTVHPSSISSLGRQLLPKTFGQSTVNISQ 60
hDP-2   ------------------------------------------------------------
mDP-2   ------------------------------------------------------------
XlDP-2  ------------------------------------------------------------
DDP     ---------------------------------------------RLQDNGLSIPKTEAGT 16
TmDP    ------------------------------------------------------------NS 2

hDP-1   QVVIGTPQRPAASNT-----LVVGS---PHTP-STHFASQNQPSDSSPWSAG-------- 103
mDP-1   QVVIGTPQRPAASNT-----IVVGS---PHTP-NTHFVSQNQTSDSSPWSAG-------- 103
XlDP-1  QVVLGTPQRQSAPNT-----ILIGS---PHTP-NTHFVSQNQATDSSPWSAG-------- 103
hDP-2   -MIISTPQRLTSSGS-----VLIGS---PYTP-APAMVTQTHIAEATGWVPGDRKRARKF 50
mDP-2   -MIISTPQRIANSGS-----VLIGN---PYTP-APAMVTQTHIAEAAGWVP--------- 41
XlDP-2  -MIISTAQRLSVAGD-----LLIGS---SYAANTSAMVTQSHITEATTWIPGDRKRAREF 51
DDP     TYTTVSAQKTSGAGSGHYDLPLKGDRYVKFTPNPIKMKSKLHAIQSNSLHSMS---ASS- 72
TmDP    ARAAMAPPRGGAAAAATAALDLTGVHILEASSVPPLPERGGNAVQRKGAVDP-------- 54
             :. :   .        : *      :.         :  :
```

FIG. 2 Amino acid alignment of TmDP protein with availablesequences of DP proteins from animal origin. *=identical amino acid.

```
hDP-1   ---------KRNRKGEKN-------GKGLRHFSMKVCEKVQRKGTTSYNEVADELVAEFSAA 149
mDP-1   ---------KRNRKGEKN-------GKGLRHFSMKVCEKVQRKGTTSYNEVADELVAEFSAA 149
X1DP-1  ---------KRNKKGEKN-------GKGLRHFSMKVCEKVQKKGTTSYNEVADELVAEFSSA 149
hDP-2   IDSDFSESKRSKKGDKN-------GKGLRHFSMKVCEKVQRKGTTSYNEVADELVSEFTNS 104
mDP-2   ---------SKRSKKGDKN-------GKGLRHFSMKVCEKVQRKGTTSYNEVADELVSEFTNS  88
X1DP-2  IESDFSESKRSKRGEKS-------GKPWRHFSMKVCEKVQTKGTTSYNEVADELVAEFTNA 105
dDP     -SS----VQRKRKPDKA-------GKGLRHFSMKVCEKVEEKGKTTYNEVADDLVSEEMK- 120
TmDP    ---------DKDRKKEKAAAPRITGWGLREYSKIVCEKVEAKGRTTYNEVADEIYSELKS- 105
                  .:.:: :*      *   *.:*  ******: ** *:*******:: :*

hDP-1   DNHILPNES-AYDQKNIRRRVYDALNVLMAMNIISKEKKEIKWIGLPTNSAQECQNLEVE 208
mDP-1   DNHILPNES-AYDQKNIRRRVYDALNVLMAMNIISKEKKEIKWIGLPTNSAQECQNLEVE 208
X1DP-1  DNHISPNESQAYDQKNIRRRVYDALNVLMAMNIISKEKKEIKWIGLPTNSAQECQNLEVE 209
hDP-2   NNHLAADS--AYDQKNIRRRVYDALNVLMAMNIISKEKKEIKWIGLPTNSAQECQNLEIE 162
mDP-2   NNHLAADSQ-AYDQENIRRRVYDALNVLMAMNIISSLPTGKKRNQVDCNSAQECQNLEIE 147
X1DP-2  AGHWPTDS--AYDQKNIRRRVYDALNVLMAMNIISKEKKEIKWIGLPPNSVQECETLEME 163
DDP     NN--AYDN--NCDQKNIRRRVYDALNVLMAINVISKDKKEIRWIGLPANSTETFLALEEE 176
TmDP    MAHIGQG----FDEKNIRRRVYDAFNVLIALRVIAKEKKEIRWMGLSNYRYEKIKKLEEV 161
                  *:::**********:***:*:.:*:. .  :  :    **

hDP-1   RQRRLERIKQKQSQLQELILQQIAFKNLVQRNRHAEQQASRPPPPNSVIHLPFIIVNTSK 268
mDP-1   RQRRLERIKQKQSQLQELILQQIAFKNLVQRNRQAEQQARRPPPPNSVIHLPFIIVNTSR 268
X1DP-1  RQRRLERIKQKQSQLQELILQQIAFKNLVQRNRLTEQKANRPPPPNSVIHLPFIIVNTSK 269
hDP-2   KQRRIERIKQKRAQLQELLLQQIAFKNLVQRNRQNEQQNQGPPALNSTIQLPFIIINTSR 222
mDP-2   KQRRIERIKQKRAQLQELLLQQIAFKNLVQRNRQNEQQNQGPPAVNSTIQLPFIIINTSR 207
X1DP-2  KQRRIERIKQKSAQLQELLVQQIAFKHLVQRNKQNEQQSRCLPAGNSTIKLPFIIVNTSK 223
DDP     NCQRRERIKQKNEMLREMIMQHVAFKGLVERNKRNESQG-VVPSPNASIQLPFIIVNTHK 235
TmDP    RKELVNKIRNKKALLQEIEKQFDDLQNIKLRNQTLESSA---ENVN-GIRLPFVLVKTSR 217
          .  . ::*:::*  *:*:  *   :: :  **:  *. .      *  *:*****:::*:
```

FIG. 2 (continued)

```
hDP-1     KTVIDCSISNDKFEYLFNFDNT-FEIHDDIEVLKRMGMACGLESGSCSAEDLKMARSLVP 327
mDP-1     KTVIDCSISNDKFEYLFNFDNT-FEIHDDIEVLKRMGMACGLESGNCSAEDLKVARSLVP 327
X1DP-1    KTVIDCSISNDKFEYLFNFDNT-FEIHDDIEVLKRMGMACGLESGSCSAEDLKTAKSLVP 328
hDP-2     KTVIDCSISSDKFEYLFNFDNT-FEIHDDIEVLKRMGMSFGLESGKCSLEDLKLAKSLVP 281
mDP-2     KTVIDCSISSDKFEYLFNFDNT-FEIHDDIEVLKRMGMSFGLESGKCSLEDLKIARSLVP 266
X1DP-2    RTVIDCSISSDKFEYLFNFDNA-FEIHDDIEVLKRMGMSFGLESGKCTSENLRLAKSFVP 282
DDP       STKINCSVTNDKSEYIFKFDKT-FEMLDDIEVLKRMGFLLGLDKGECTPENIERVKSWVP 294
TmDP      KARVEIEISDDSKFAHFEFNGAPFTLHDDLSILEGVRGNSIGKAGRATLH---------- 267
                : :: .::.*.    *:*: : * : **:.:*: :     . * .: .

hDP-1     KALEPYVTEMAQGT--VGGVFITT-AGSTSNGTRFS----------ASDLTNGADGMLAT- 374
mDP-1     KALEPYVTEMAQGS--IGGVFVTT-TGSTSNGTRLS----------ASDLSNGADGMLAT- 374
X1DP-1    KALEPYVTEMAQGS--ISSVYISFSSGSVSNGRRFS----------SSDLTGCTDGMLAT- 376
hDP-2     KALEGYITDISTGPSWLNQGLLLNSTQSVSNLDLTTGATLPQSSVNQGLCLDAEVALATG 341
mDP-2     KALEGYITDISTGPSWLNQGLLLNSTQSVSNLDPTTGATVPQSSVNQGLCLDAEVALATG 326
X1DP-2    RALEGYVTEMATGSSWADQSFNSQTVSSAGNTTCTT--TNSQASLTSGLYFDSEVSLTT- 339
DDP       PNLAKYVEAYGTGKT-GENMYESDDEDNEFNGYLES----------ANESQGFAQH- 339
TmDP      ---------------------------------------------------------

hDP-1     ----------SSNGSQYSGSRVETPVSYVGEDDEEDDDFNENDEDD 410
mDP-1     ----------SSNGSQYSGSRVETPVSYVGEDDDDDDDFNENDEED 410
X1DP-1    ----------SSNGSQYSSSRVETPVSYVEEEDDDDDDDLDDD--- 409
hDP-2     QFLAPNSHQSSSAASHCSESRGETPCSFNDEDEEDDEEDSSSPE-- 385
mDP-2     QLPASNSHQSSSAASHFSESRGETPCSFNDEDEEDEEEDPSSPE-- 370
X1DP-2    -----SSHHSSSGTSHYTESRGETPCWFDDD--EDDEDDSSSME-- 376
dDP       ----------S-AQHTTD--GEFKLEMDDDELDDDID-------- 363
TmDP      ---------------------------------------------
```

FIG. 2 (continued)

```
 Score = 79 (27.8 bits), Expect = 9.5, P = 1.0
 Identities = 46/188 (24%), Positives = 98/188 (52%), Frame = +3

Query:    64 GW--GLREYSKIVCEKVEAKGRTTYNEVADEIYSELKSMAHIGQGFDEKNIR-RRVYDAF120
             GW    R  S+ V  +VEA G +   +  +E+   +K  ++   GF EK++   ++ + +
Sbjct: 75510 GWYNESRRDSETVKARVEA-GLSEVKKSVEELALLIKR-SNRSAGFQEKDMEVLKMEEKY75683

Query:   121 -NVLIALRVIAKEKKEIRWMGLSNYRYEKI---KKLEEVRKELVNKIRNKKALLQEIEKQ176
               V+  L V+ +E   ++ + +S+   E++   +K+EE+R+  +    +R  ++L +EIE
Sbjct: 75684 AEVMRVLEVVKEEVSRVK-LDVSSVLIERVAAEEKVEELRFKTEGGLRLLESLKKEIEVA75860

Query:   177 FDD-----LQNIK-LRN-QTLESSAENVNGIRLPFVLVKTSRKARVEIEISDDSKFAHFE229
              ++     L  I+ L+  + +E  E    I++  +LV+ +++ +  +E ++ SK    E
Sbjct: 75861 NEEHLMVALGKIEALKGYKEIERQREG-KAIKVLDLLVERNKRIKNMLEEAERSKDIEIE76037

Query:   230 FNGAPFTLHDDLSILE 245
                  F    D+ +LE
Sbjct: 76038 L----FETSTDVEMLE 76073
```

FIG. 3

SEQUENCE LISTING

<110> CONSEJO SUPERIOR DE INVESTICIONES CIENTIFICAS
GUTIERREZ-ARMENTA, CRISANTO
RAMIREZ-PARRA, ELENA

<120> WHEAT DP PROTEINS AND USES THEREOF

<130> 141183

<140>
<141>

<150> ES9902127
<151> 1999-09-24

<150> ES9902474
<151> 1999-11-11

<160> 16

<170> PatentIn Ver. 2.1

<210> 1
<211> 1089
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (20)..(802)

<400> 1
gaattcggca cgagccgca atg gcg cct ccc cgc ggc gga gct gct gcg gcc 52
Met Ala Pro Pro Arg Gly Gly Ala Ala Ala Ala
1 5 10

gct acc gcc gca ctg gac ctg acc ggc gtg cac att ctc gaa gct tcc 100
Ala Thr Ala Ala Leu Asp Leu Thr Gly Val His Ile Leu Glu Ala Ser
15 20 25

agt gtc ccc ccg ctt ccc gaa cgc ggc ggt aat gcg gtc caa agg aag 148
Ser Val Pro Pro Leu Pro Glu Arg Gly Gly Asn Ala Val Gln Arg Lys
30 35 40

ggg gct gtt gac ccg gat aaa gat agg aag aag gag aag gct gcg gca 196
Gly Ala Val Asp Pro Asp Lys Asp Arg Lys Lys Glu Lys Ala Ala Ala
45 50 55

```
ccg agg atc acc ggt tgg ggg ctc cgc gag tac agc aaa ata gtt tgt   244
Pro Arg Ile Thr Gly Trp Gly Leu Arg Glu Tyr Ser Lys Ile Val Cys
 60                  65                  70                  75

gag aaa gtt gaa gcc aaa gga aga aca aca tac aat gag gtt gca gac   292
Glu Lys Val Glu Ala Lys Gly Arg Thr Thr Tyr Asn Glu Val Ala Asp
                 80                  85                  90

gaa att tat tca gag ctg aag tcc atg gca cat att ggt caa ggg ttt   340
Glu Ile Tyr Ser Glu Leu Lys Ser Met Ala His Ile Gly Gln Gly Phe
             95                 100                 105

gat gag aag aat att agg cgg aga gtg tat gat gct ttc aac gtt ctc   388
Asp Glu Lys Asn Ile Arg Arg Arg Val Tyr Asp Ala Phe Asn Val Leu
        110                 115                 120

att gca ctt cgt gtt att gca aaa gaa aaa aag gag ata cgg tgg atg   436
Ile Ala Leu Arg Val Ile Ala Lys Glu Lys Lys Glu Ile Arg Trp Met
    125                 130                 135

ggc ctt tca aat tac aga tat gaa aaa ata aag aag ctt gag gaa gtt   484
Gly Leu Ser Asn Tyr Arg Tyr Glu Lys Ile Lys Lys Leu Glu Glu Val
140                 145                 150                 155

cgt aaa gaa ctc gtc aac aag att agg aac aag aag gca ctc ctc cag   532
Arg Lys Glu Leu Val Asn Lys Ile Arg Asn Lys Lys Ala Leu Leu Gln
                160                 165                 170

gaa atc gaa aaa cag ttt gat gat ctc caa aac atc aag tta cgt aac   580
Glu Ile Glu Lys Gln Phe Asp Asp Leu Gln Asn Ile Lys Leu Arg Asn
            175                 180                 185

caa aca ctg gaa agc tca gca gag aat gtt aat ggc atc cgc ctt cca   628
Gln Thr Leu Glu Ser Ser Ala Glu Asn Val Asn Gly Ile Arg Leu Pro
        190                 195                 200

ttc gta ttg gtc aag aca tct agg aaa gca agg gtg gaa att gag att   676
Phe Val Leu Val Lys Thr Ser Arg Lys Ala Arg Val Glu Ile Glu Ile
    205                 210                 215

tca gat gac tcg aag ttt gcc cat ttc gag ttc aat ggt gca cca ttc   724
Ser Asp Asp Ser Lys Phe Ala His Phe Glu Phe Asn Gly Ala Pro Phe
220                 225                 230                 235

aca ttg cat gat gat ctc tca atc ctt gag ggg gta agg cgt aac agc   772
Thr Leu His Asp Asp Leu Ser Ile Leu Glu Gly Val Arg Arg Asn Ser
                240                 245                 250
```

```
ata gga aga gct ggc cgc gcc acc ctt cac tagagactca agaatattac     822
Ile Gly Arg Ala Gly Arg Ala Thr Leu His
            255                 260

aaatgaatta aaagtgttag aactggcaca gccggattct tttgcacagc tatgtatagc 882

tatatatcct catgaaaact tgacctagtt tataggacag tctctcaggc ttgagaagat 942

tttaacctgc aaattttgtc tccttttttgt gcctagcagg ttattaggtc tcagatagat 1002

gattcatata tgtgctgcta tgaaaacatt gatagcaaaa aaaaaaaaaa aaaaaaaaaa 1062

aaaaaaaaaa aaaaaaaaaa aaaaaaa                                     1089
```

<210> 2
<211> 261
<212> PRT
<213> Triticum monococcum

<400> 2

```
Met Ala Pro Pro Arg Gly Gly Ala Ala Ala Ala Ala Thr Ala Ala Leu
  1               5                  10                  15

Asp Leu Thr Gly Val His Ile Leu Glu Ala Ser Ser Val Pro Pro Leu
             20                  25                  30

Pro Glu Arg Gly Gly Asn Ala Val Gln Arg Lys Gly Ala Val Asp Pro
         35                  40                  45

Asp Lys Asp Arg Lys Lys Glu Lys Ala Ala Ala Pro Arg Ile Thr Gly
     50                  55                  60

Trp Gly Leu Arg Glu Tyr Ser Lys Ile Val Cys Glu Lys Val Glu Ala
 65                  70                  75                  80

Lys Gly Arg Thr Thr Tyr Asn Glu Val Ala Asp Glu Ile Tyr Ser Glu
                 85                  90                  95

Leu Lys Ser Met Ala His Ile Gly Gln Gly Phe Asp Glu Lys Asn Ile
            100                 105                 110

Arg Arg Arg Val Tyr Asp Ala Phe Asn Val Leu Ile Ala Leu Arg Val
        115                 120                 125

Ile Ala Lys Glu Lys Lys Glu Ile Arg Trp Met Gly Leu Ser Asn Tyr
    130                 135                 140
```

```
Arg Tyr Glu Lys Ile Lys Lys Leu Glu Glu Val Arg Lys Glu Leu Val
145                 150                 155                 160

Asn Lys Ile Arg Asn Lys Lys Ala Leu Leu Gln Glu Ile Glu Lys Gln
                165                 170                 175

Phe Asp Asp Leu Gln Asn Ile Lys Leu Arg Asn Gln Thr Leu Glu Ser
            180                 185                 190

Ser Ala Glu Asn Val Asn Gly Ile Arg Leu Pro Phe Val Leu Val Lys
        195                 200                 205

Thr Ser Arg Lys Ala Arg Val Glu Ile Glu Ile Ser Asp Asp Ser Lys
    210                 215                 220

Phe Ala His Phe Glu Phe Asn Gly Ala Pro Phe Thr Leu His Asp Asp
225                 230                 235                 240

Leu Ser Ile Leu Glu Gly Val Arg Arg Asn Ser Ile Gly Arg Ala Gly
                245                 250                 255

Arg Ala Thr Leu His
            260
```

<210> 3
<211> 1974
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (166)..(1539)

<400> 3
aattcggcac gagcccaccc acctacctcc cgccgccgcc gccgccacgg gaaccctatc 60

tccggcgagc ccccccccgat gccctcctgc cttctctgaa gccgaagacg cccatcgctc 120

ccgggagtcg ggggtcccgc agcgcgcgat cgcgagatcg ggctt atg tct ggg ggc 177
                                                    Met Ser Gly Gly
                                                      1

```
ggc agg ccg ccg gct gcg caa aaa atc ctg cag tct ctg cgc ccg ccc   225
Gly Arg Pro Pro Ala Ala Gln Lys Ile Leu Gln Ser Leu Arg Pro Pro
  5                  10                  15                  20
```

```
ccg gtg ttc tcc acg ccg tcg cgg cct ccc ttc gcc tca ccc gac gac   273
Pro Val Phe Ser Thr Pro Ser Arg Pro Pro Phe Ala Ser Pro Asp Asp
                 25                  30                  35

tac cac cgc ttt cat gcg ccg act acc cct tct gcc act ggc tcc ggc   321
Tyr His Arg Phe His Ala Pro Thr Thr Pro Ser Ala Thr Gly Ser Gly
             40                  45                  50

ggc atc ggc tcc ggt ggt gtt ggc ggc gat att gat gag ggg ctt gtt   369
Gly Ile Gly Ser Gly Gly Val Gly Gly Asp Ile Asp Glu Gly Leu Val
         55                  60                  65

atc cgg acg cag cta aaa aga aaa gcc aca cgc gaa gaa aat aat gcg   417
Ile Arg Thr Gln Leu Lys Arg Lys Ala Thr Arg Glu Glu Asn Asn Ala
     70                  75                  80

gct gag tcg agt gac tgt atg att gtc acc act gga gtt act ggc aat   465
Ala Glu Ser Ser Asp Cys Met Ile Val Thr Thr Gly Val Thr Gly Asn
 85                  90                  95                 100

ccg cta ctc acc cca gtg tct gga aaa gct gtt aag aat tct aaa tca   513
Pro Leu Leu Thr Pro Val Ser Gly Lys Ala Val Lys Asn Ser Lys Ser
                105                 110                 115

aag act aag aac aat aaa gct ggg cct cag aca cct acg cca aat gtt   561
Lys Thr Lys Asn Asn Lys Ala Gly Pro Gln Thr Pro Thr Pro Asn Val
            120                 125                 130

ggc tca cca ctc aat cca tca act cct gct ggt act tgc cgc tat gac   609
Gly Ser Pro Leu Asn Pro Ser Thr Pro Ala Gly Thr Cys Arg Tyr Asp
        135                 140                 145

agt tcg tta gga ctt ctg aca aag aag ttc atc aat ttg ctg aag caa   657
Ser Ser Leu Gly Leu Leu Thr Lys Lys Phe Ile Asn Leu Leu Lys Gln
    150                 155                 160

gct gag gat ggc att cta gat ttg aat aat gct gca gaa aca cta gag   705
Ala Glu Asp Gly Ile Leu Asp Leu Asn Asn Ala Ala Glu Thr Leu Glu
165                 170                 175                 180

gtt caa aag cga cgc ata tat gac atc aca aat gtc ctc gaa gga att   753
Val Gln Lys Arg Arg Ile Tyr Asp Ile Thr Asn Val Leu Glu Gly Ile
                185                 190                 195

ggt ctt ata gaa aag aca ctt aaa aac aga att cgt tgg aag ggc ttg   801
Gly Leu Ile Glu Lys Thr Leu Lys Asn Arg Ile Arg Trp Lys Gly Leu
            200                 205                 210
```

```
gat gat tca gga gtg gaa tta gat aat ggc ctt tca ggt ttg cag aca   849
Asp Asp Ser Gly Val Glu Leu Asp Asn Gly Leu Ser Gly Leu Gln Thr
        215                 220                 225

gaa gtt gaa aat ctt aat ttg cag gag caa gcc tta gat gag cgt ata   897
Glu Val Glu Asn Leu Asn Leu Gln Glu Gln Ala Leu Asp Glu Arg Ile
    230                 235                 240

agt gat atg cgc gaa aag cta agg ggg tta acg gaa gat gag aac agt   945
Ser Asp Met Arg Glu Lys Leu Arg Gly Leu Thr Glu Asp Glu Asn Ser
245                 250                 255                 260

caa aga tgg ctc tat gtg acg gaa gat gat atc aag gga tta ccc tgc   993
Gln Arg Trp Leu Tyr Val Thr Glu Asp Asp Ile Lys Gly Leu Pro Cys
                265                 270                 275

ttt cag aat gaa act cta att gca ata aaa gct cct cat ggt act aca   1041
Phe Gln Asn Glu Thr Leu Ile Ala Ile Lys Ala Pro His Gly Thr Thr
            280                 285                 290

ctt gaa gta cct gat cct gat gag gct ggt gat tat ctc cag agg aga   1089
Leu Glu Val Pro Asp Pro Asp Glu Ala Gly Asp Tyr Leu Gln Arg Arg
        295                 300                 305

tac aga atc gta tta aga agt acc ctg ggt cca ata gat gtt tac tta   1137
Tyr Arg Ile Val Leu Arg Ser Thr Leu Gly Pro Ile Asp Val Tyr Leu
    310                 315                 320

gtt agt caa ttt gac gat gga ttt gag aat ttg ggt ggt gct gcg aca   1185
Val Ser Gln Phe Asp Asp Gly Phe Glu Asn Leu Gly Gly Ala Ala Thr
325                 330                 335                 340

cct cca agg cat aca aat gtc cca aaa cct gga cct tgt gaa gac tta   1233
Pro Pro Arg His Thr Asn Val Pro Lys Pro Gly Pro Cys Glu Asp Leu
                345                 350                 355

cat gca aca aac gct aca caa agc agc aaa tca atc aat gtg gaa tat   1281
His Ala Thr Asn Ala Thr Gln Ser Ser Lys Ser Ile Asn Val Glu Tyr
            360                 365                 370

aat att cag cac agg cag aat act cca caa gat cct agt tct tca aat   1329
Asn Ile Gln His Arg Gln Asn Thr Pro Gln Asp Pro Ser Ser Ser Asn
        375                 380                 385

gat tat gga ggg atg aca agg ata atc cct tca gat gtt aat act gat   1377
Asp Tyr Gly Gly Met Thr Arg Ile Ile Pro Ser Asp Val Asn Thr Asp
    390                 395                 400
```

```
gct gat tac tgg ctc cta aca gag ggt gat gtt agt att act gac atg   1425
Ala Asp Tyr Trp Leu Leu Thr Glu Gly Asp Val Ser Ile Thr Asp Met
405                 410                 415                 420

tgg gaa aca gca cca gaa gtg cag tgg gac acc gct gtg ttt tta cct   1473
Trp Glu Thr Ala Pro Glu Val Gln Trp Asp Thr Ala Val Phe Leu Pro
                425                 430                 435

gaa gat gtt agc atc cca cat gca cat cat agt ccg cgg atg cag gtt   1521
Glu Asp Val Ser Ile Pro His Ala His His Ser Pro Arg Met Gln Val
            440                 445                 450

cca agc atg gat caa cca taaggtcatg gcggtgaaaa cttgacatat          1569
Pro Ser Met Asp Gln Pro
        455

ggaattcctg gagtgctgtt tcagaaaata ctgatttcaa aatggaaaga tcagggcagc 1629

aagttcagac tgatcaccgt tctgaatttg ctgtttgtta tggagacgat tggtgccaac 1689

taacttatca gtctgctgcc ttgtttgttc tggcacctgt ccttcagttg aaaaggcgcc 1749

catgtgcata ttgcaccttg aattcgggct gctatgcaca ttcggtatct gctttatttc 1809

tctaactgag tatattttgc aaggcaatag tggctctgta gctctcttgg gaattaatac 1869

gaatcttttt gagcaaaaac agtagggaag tcccctgttg tgactctttc attatataaa 1929

tggagtttat acaaaggggt aaaaaaaaaa aaaaaaaaaa aaaaa                 1974
```

<210> 4
<211> 458
<212> PRT
<213> Triticum monococcum

<400> 4

```
Met Ser Gly Gly Gly Arg Pro Pro Ala Ala Gln Lys Ile Leu Gln Ser
  1               5                  10                  15

Leu Arg Pro Pro Pro Val Phe Ser Thr Pro Ser Arg Pro Pro Phe Ala
             20                  25                  30

Ser Pro Asp Asp Tyr His Arg Phe His Ala Pro Thr Thr Pro Ser Ala
         35                  40                  45

Thr Gly Ser Gly Gly Ile Gly Ser Gly Gly Val Gly Gly Asp Ile Asp
```

```
     50                  55                  60

Glu Gly Leu Val Ile Arg Thr Gln Leu Lys Arg Lys Ala Thr Arg Glu
 65                  70                  75                  80

Glu Asn Asn Ala Ala Glu Ser Ser Asp Cys Met Ile Val Thr Thr Gly
                 85                  90                  95

Val Thr Gly Asn Pro Leu Leu Thr Pro Val Ser Gly Lys Ala Val Lys
            100                 105                 110

Asn Ser Lys Ser Lys Thr Lys Asn Asn Lys Ala Gly Pro Gln Thr Pro
        115                 120                 125

Thr Pro Asn Val Gly Ser Pro Leu Asn Pro Ser Thr Pro Ala Gly Thr
    130                 135                 140

Cys Arg Tyr Asp Ser Ser Leu Gly Leu Leu Thr Lys Lys Phe Ile Asn
145                 150                 155                 160

Leu Leu Lys Gln Ala Glu Asp Gly Ile Leu Asp Leu Asn Asn Ala Ala
                165                 170                 175

Glu Thr Leu Glu Val Gln Lys Arg Arg Ile Tyr Asp Ile Thr Asn Val
            180                 185                 190

Leu Glu Gly Ile Gly Leu Ile Glu Lys Thr Leu Lys Asn Arg Ile Arg
        195                 200                 205

Trp Lys Gly Leu Asp Asp Ser Gly Val Glu Leu Asp Asn Gly Leu Ser
    210                 215                 220

Gly Leu Gln Thr Glu Val Glu Asn Leu Asn Leu Gln Glu Gln Ala Leu
225                 230                 235                 240

Asp Glu Arg Ile Ser Asp Met Arg Glu Lys Leu Arg Gly Leu Thr Glu
                245                 250                 255

Asp Glu Asn Ser Gln Arg Trp Leu Tyr Val Thr Glu Asp Asp Ile Lys
            260                 265                 270

Gly Leu Pro Cys Phe Gln Asn Glu Thr Leu Ile Ala Ile Lys Ala Pro
        275                 280                 285

His Gly Thr Thr Leu Glu Val Pro Asp Pro Asp Glu Ala Gly Asp Tyr
    290                 295                 300

Leu Gln Arg Arg Tyr Arg Ile Val Leu Arg Ser Thr Leu Gly Pro Ile
```

```
305                 310                 315                 320

Asp Val Tyr Leu Val Ser Gln Phe Asp Asp Gly Phe Glu Asn Leu Gly
                325                 330                 335

Gly Ala Ala Thr Pro Pro Arg His Thr Asn Val Pro Lys Pro Gly Pro
            340                 345                 350

Cys Glu Asp Leu His Ala Thr Asn Ala Thr Gln Ser Ser Lys Ser Ile
        355                 360                 365

Asn Val Glu Tyr Asn Ile Gln His Arg Gln Asn Thr Pro Gln Asp Pro
    370                 375                 380

Ser Ser Ser Asn Asp Tyr Gly Gly Met Thr Arg Ile Ile Pro Ser Asp
385                 390                 395                 400

Val Asn Thr Asp Ala Asp Tyr Trp Leu Leu Thr Glu Gly Asp Val Ser
                405                 410                 415

Ile Thr Asp Met Trp Glu Thr Ala Pro Glu Val Gln Trp Asp Thr Ala
            420                 425                 430

Val Phe Leu Pro Glu Asp Val Ser Ile Pro His Ala His His Ser Pro
        435                 440                 445

Arg Met Gln Val Pro Ser Met Asp Gln Pro
    450                 455
```

<210> 5
<211> 156
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (1)..(156)

<400> 5

```
gca cga gcc gca atg gcg cct ccc cgc ggc gga gct gct gcg gcc gct   48
Ala Arg Ala Ala Met Ala Pro Pro Arg Gly Gly Ala Ala Ala Ala Ala
  1               5                  10                  15

acc gcc gca ctg gac ctg acc ggc gtg cac att ctc gaa gct tcc agt   96
Thr Ala Ala Leu Asp Leu Thr Gly Val His Ile Leu Glu Ala Ser Ser
            20                  25                  30
```

```
gtc ccc ccg ctt ccc gaa cgc ggc ggt aat gcg gtc caa agg aag ggg   144
Val Pro Pro Leu Pro Glu Arg Gly Gly Asn Ala Val Gln Arg Lys Gly
         35                  40                  45

gct gtt gac ccg                                                  156
Ala Val Asp Pro
     50
```

<210> 6
<211> 52
<212> PRT
<213> Triticum monococcum

<400> 6

```
Ala Arg Ala Ala Met Ala Pro Pro Arg Gly Gly Ala Ala Ala Ala Ala
  1               5                  10                  15

Thr Ala Ala Leu Asp Leu Thr Gly Val His Ile Leu Glu Ala Ser Ser
             20                  25                  30

Val Pro Pro Leu Pro Glu Arg Gly Gly Asn Ala Val Gln Arg Lys Gly
         35                  40                  45

Ala Val Asp Pro
     50
```

<210> 7
<211> 153
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (1)..(153)

<400> 7

```
gat aaa gat agg aag aag gag aag gct gcg gca ccg agg atc acc ggt   48
Asp Lys Asp Arg Lys Lys Glu Lys Ala Ala Ala Pro Arg Ile Thr Gly
  1               5                  10                  15

tgg ggg ctc cgc gag tac agc aaa ata gtt tgt gag aaa gtt gaa gcc   96
Trp Gly Leu Arg Glu Tyr Ser Lys Ile Val Cys Glu Lys Val Glu Ala
             20                  25                  30
```

```
aaa gga aga aca aca tac aat gag gtt gca gac gaa att tat tca gag   144
Lys Gly Arg Thr Thr Tyr Asn Glu Val Ala Asp Glu Ile Tyr Ser Glu
         35                  40                  45

ctg aag tcc                                                      153
Leu Lys Ser
     50
```

<210> 8
<211> 51
<212> PRT
<213> Triticum monococcum

<400> 8

```
Asp Lys Asp Arg Lys Lys Glu Lys Ala Ala Ala Pro Arg Ile Thr Gly
  1               5                  10                  15

Trp Gly Leu Arg Glu Tyr Ser Lys Ile Val Cys Glu Lys Val Glu Ala
             20                  25                  30

Lys Gly Arg Thr Thr Tyr Asn Glu Val Ala Asp Glu Ile Tyr Ser Glu
         35                  40                  45

Leu Lys Ser
     50
```

<210> 9
<211> 168
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (1)..(168)

<400> 9

```
atg gca cat att ggt caa ggg ttt gat gag aag aat att agg cgg aga   48
Met Ala His Ile Gly Gln Gly Phe Asp Glu Lys Asn Ile Arg Arg Arg
  1               5                  10                  15

gtg tat gat gct ttc aac gtt ctc att gca ctt cgt gtt att gca aaa   96
Val Tyr Asp Ala Phe Asn Val Leu Ile Ala Leu Arg Val Ile Ala Lys
             20                  25                  30

gaa aaa aag gag ata cgg tgg atg ggc ctt tca aat tac aga tat gaa   144
```

```
Glu Lys Lys Glu Ile Arg Trp Met Gly Leu Ser Asn Tyr Arg Tyr Glu
         35                  40                  45

aaa ata aag aag ctt gag gaa gtt                                    168
Lys Ile Lys Lys Leu Glu Glu Val
     50                  55
```

<210> 10
<211> 56
<212> PRT
<213> Triticum monococcum

<400> 10

```
Met Ala His Ile Gly Gln Gly Phe Asp Glu Lys Asn Ile Arg Arg Arg
  1               5                  10                  15

Val Tyr Asp Ala Phe Asn Val Leu Ile Ala Leu Arg Val Ile Ala Lys
             20                  25                  30

Glu Lys Lys Glu Ile Arg Trp Met Gly Leu Ser Asn Tyr Arg Tyr Glu
         35                  40                  45

Lys Ile Lys Lys Leu Glu Glu Val
     50                  55
```

<210> 11
<211> 168
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (1)..(168)

<400> 11

```
cgt aaa gaa ctc gtc aac aag att agg aac aag aag gca ctc ctc cag   48
Arg Lys Glu Leu Val Asn Lys Ile Arg Asn Lys Lys Ala Leu Leu Gln
  1               5                  10                  15

gaa atc gaa aaa cag ttt gat gat ctc caa aac atc aag tta cgt aac   96
Glu Ile Glu Lys Gln Phe Asp Asp Leu Gln Asn Ile Lys Leu Arg Asn
             20                  25                  30

caa aca ctg gaa agc tca gca gag aat gtt aat ggc atc cgc ctt cca   144
Gln Thr Leu Glu Ser Ser Ala Glu Asn Val Asn Gly Ile Arg Leu Pro
```

```
         35                  40                  45

ttc gta ttg gtc aag aca tct agg                                    168
Phe Val Leu Val Lys Thr Ser Arg
     50                  55
```

<210> 12
<211> 56
<212> PRT
<213> Triticum monococcum

<400> 12

```
Arg Lys Glu Leu Val Asn Lys Ile Arg Asn Lys Lys Ala Leu Leu Gln
  1               5                  10                  15

Glu Ile Glu Lys Gln Phe Asp Asp Leu Gln Asn Ile Lys Leu Arg Asn
             20                  25                  30

Gln Thr Leu Glu Ser Ser Ala Glu Asn Val Asn Gly Ile Arg Leu Pro
         35                  40                  45

Phe Val Leu Val Lys Thr Ser Arg
     50                  55
```

<210> 13
<211> 150
<212> DNA
<213> Triticum monococcum

<220>
<221> CDS
<222> (1)..(150)

<400> 13

```
aaa gca agg gtg gaa att gag att tca gat gac tcg aag ttt gcc cat   48
Lys Ala Arg Val Glu Ile Glu Ile Ser Asp Asp Ser Lys Phe Ala His
  1               5                  10                  15

ttc gag ttc aat ggt gca cca ttc aca ttg cat gat gat ctc tca atc   96
Phe Glu Phe Asn Gly Ala Pro Phe Thr Leu His Asp Asp Leu Ser Ile
             20                  25                  30

ctt gag ggg gta agg cgt aac agc ata gga aga gct ggc cgc gcc acc   144
Leu Glu Gly Val Arg Arg Asn Ser Ile Gly Arg Ala Gly Arg Ala Thr
         35                  40                  45
```

ctt cac 150
Leu His
50

<210> 14
<211> 50
<212> PRT
<213> Triticum monococcum

<400> 14
Lys Ala Arg Val Glu Ile Glu Ile Ser Asp Asp Ser Lys Phe Ala His
1 5 10 15

Phe Glu Phe Asn Gly Ala Pro Phe Thr Leu His Asp Asp Leu Ser Ile
20 25 30

Leu Glu Gly Val Arg Arg Asn Ser Ile Gly Arg Ala Gly Arg Ala Thr
35 40 45

Leu His
50

<210> 15
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_binding
<222> (8)..(16)

<220>
<223> Description of Artificial Sequence:E2F canonical binding site probe

<400> 15
atttaagttt cgcgcccttt ctcaa 25

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_binding
<222> (8)..(16)
<223> Mutant E2F Canonical binding site probe

<220>
<223> Description of Artificial Sequence:Mutant E2F
Canonical binding site probe

<400> 16
atttaagttt cgatcccttt ctcaa 25